The Order of the Giant Component of Random Hypergraphs

Michael Behrisch*, Amin Coja-Oghlan**, and Mihyun Kang

Humboldt-Universität zu Berlin, Institut für Informatik,
Unter den Linden 6, 10099 Berlin, Germany
coja@informatik.hu-berlin.de

Abstract. We establish central and local limit theorems for the number of vertices in the largest component of a random d-uniform hypergraph $H_d(n,p)$ with edge probability $p = c/\binom{n-1}{d-1}$, where $(d-1)^{-1} + \varepsilon < c < \infty$. The proof relies on a new, purely probabilistic approach, and is based on Stein's method as well as exposing the edges of $H_d(n,p)$ in several rounds.
Keywords: random graphs and hypergraphs, limit theorems, giant component, Stein's method.

1 Introduction and Results

A d*-uniform hypergraph* $H = (V,E)$ consists of a set V of vertices and a set E of edges, which are subsets of V of cardinality d. Moreover, a vertex w is *reachable in* H from a vertex v if either $v = w$ or there is a sequence $e_1,\ldots,e_k$ of edges such that $v \in e_1$, $w \in e_k$, and $e_i \cap e_{i+1} \neq \emptyset$ for $i = 1,\ldots,k-1$. Of course, reachability in H is an equivalence relation. The equivalence classes are the *components* of H, and H is *connected* if there is only one component.

Throughout the paper, we let $V = \{1,\ldots,n\}$ be a set of n vertices. Moreover, if $2 \leq d$ is a fixed integer and $0 \leq p = p(n) \leq 1$ is sequence, then we let $H_d(n,p)$ signify a random d-uniform hypergraph with vertex set V in which each of the $\binom{n}{d}$ possible edges is present with probability p independently. We say that $H_d(n,p)$ enjoys some property $\mathcal{P}$ *with high probability* (w.h.p.) if the probability that $H_d(n,p)$ has $\mathcal{P}$ tends to 1 as $n \to \infty$. If $d = 2$, then the $H_d(n,p)$ model is identical with the well-known $G(n,p)$ model of random graphs. In order to state some related results we will also need a different model $H_d(n,m)$ of random hypergraphs, where the hypergraph is chosen uniformly at random among all d-uniform hypergraphs with n vertices and m edges.

Since the pioneering work of Erdős and Rényi [8], the component structure of random discrete structures has been a central theme in probabilistic combinatorics. In the present paper, we contribute to this theme by analyzing the maximum order $\mathcal{N}(H_d(n,p))$ of a component of $H_d(n,p)$ in greater detail. More precisely, establishing central and local limit theorems for $\mathcal{N}(H_d(n,p))$, we determine the asymptotic distribution of $\mathcal{N}(H_d(n,p))$ precisely. Though such limit theorems are known in the case of graphs (i.e, $d = 2$), they are new in the case of d-uniform hypergraphs for $d > 2$. Indeed, to the best of our knowledge none of the arguments known for the graph case extends directly to the case of hypergraphs ($d > 2$). Therefore, we present a new, purely probabilistic proof of the central and local limit theorems, which, in contrast to most prior work, does not rely on involved enumerative techniques. We believe that this new technique is interesting in its own right and may have further applications.

The giant component. In their seminal paper [8], Erdős and Rényi proved that the number of vertices in the largest component of $G(n,p)$ undergoes a *phase transition* as $np \sim 1$. They showed that if $np < 1 - \varepsilon$ for an arbitrarily small $\varepsilon > 0$ that remains fixed as $n \to \infty$, then all components of $G(n,p)$ consist of $O(\ln n)$ vertices. By contrast, if $np > 1 + \varepsilon$, then $G(n,p)$ has one *giant* component on a linear number $\Omega(n)$ of vertices, while all other components contain only $O(\ln n)$ vertices. In fact, in the case $1+\varepsilon < c = (n-1)p = O(1)$ Erdős and Rényi estimated the order (i.e., the number of vertices) of the giant component: let $\mathcal{N}(G(n,p))$ signify the maximum order of a component of $G(n,p)$. Then

$$n^{-1}\mathcal{N}(G(n,p)) \text{ converges in distribution to the constant } 1-\rho, \tag{1}$$

* Supported by the DFG research center MATHEON in Berlin.
** Supported by the Deutsche Forschungsgemeinschaft (DFG FOR 413/1-2)

where $0 < \rho < 1$ is the unique solution to the transcendental equation $\rho = \exp(c(\rho - 1))$.

A corresponding result was established by Schmidt-Pruzan and Shamir [17] for random hypergraphs $H_d(n,p)$. They showed that a random hypergraph $H_d(n,p)$ consists of components of order $O(\ln n)$ if $(d-1)\binom{n-1}{d-1}p < 1-\varepsilon$, whereas $H_d(n,p)$ has a unique large (the *giant*) component on $\Omega(n)$ vertices w.h.p. if $(d-1)\binom{n-1}{d-1}p > 1+\varepsilon$. Furthermore, Coja-Oghlan, Moore, and Sanwalani [7] established a result similar to (1), showing that in the case $(d-1)\binom{n-1}{d-1}p > 1+\varepsilon$ the order of the giant component is $(1-\rho)n + o(n)$ w.h.p., where $0 < \rho < 1$ is the unique solution to the transcendental equation

$$\rho = \exp(c(\rho^{d-1} - 1)). \tag{2}$$

Central and local limit theorems. In terms of limit theorems, (1) provides a *strong law of large numbers* for $\mathcal{N}(G(n,p))$, i.e., it yields the probable value of $\mathcal{N}(G(n,p))$ up to fluctuations of order $o(n)$. Thus, a natural question is whether we can characterize the distribution of $\mathcal{N}(G(n,p))$ (or $\mathcal{N}(H_d(n,p))$) more precisely; for instance, is it true that $\mathcal{N}(G(n,p))$ "converges to the normal distribution" in some sense? Our first result, which we will prove in Section 5, shows that this is indeed the case.

Theorem 1. *Let $\mathcal{J} \subset ((d-1)^{-1}, \infty)$ be a compact interval, and let $0 \le p = p(n) \le 1$ be a sequence such that $c = c(n) = \binom{n-1}{d-1}p \in \mathcal{J}$ for all n. Furthermore, let $0 < \rho = \rho(n) < 1$ be the unique solution to (2), and set*

$$\sigma^2 = \sigma(n)^2 = \frac{\rho\left[1 - \rho + c(d-1)(\rho - \rho^{d-1})\right] n}{(1 - c(d-1)\rho^{d-1})^2}. \tag{3}$$

Then $\sigma^{-1}(\mathcal{N}(H_d(n,p)) - (1-\rho)n)$ converges in distribution to the standard normal distribution.

Theorem 1 provides a *central limit theorem* for $\mathcal{N}(H_d(n,p))$; it shows that for any fixed numbers $a < b$

$$\lim_{n\to\infty} \mathrm{P}\left[a \le \frac{\mathcal{N}(H_d(n,p)) - (1-\rho)n}{\sigma} \le b\right] = (2\pi)^{-\frac{1}{2}} \int_a^b \exp(-t^2/2)dt \tag{4}$$

(provided that the sequence $p = p(n)$ satisfies the above assumptions).

Though Theorem 1 provides quite useful information about the distribution of $\mathcal{N}(H_d(n,p))$, the main result of this paper is actually a *local limit theorem* for $\mathcal{N}(H_d(n,p))$, which characterizes the distribution of $\mathcal{N}(H_d(n,p))$ even more precisely. To motivate the local limit theorem, we emphasize that Theorem 1 only estimates $\mathcal{N}(G(n,p))$ up to an error of $o(\sigma) = o(\sqrt{n})$. That is, we do obtain from (4) that for arbitrarily small but fixed $\gamma > 0$

$$\mathrm{P}\left[|\mathcal{N}(H_d(n,p)) - \nu| \le \gamma\sigma\right] \sim \frac{1}{\sqrt{2\pi}\sigma} \int_{-\gamma\sigma}^{\gamma\sigma} \exp\left[\frac{(\nu - (1-\rho)n - t)^2}{2\sigma^2}\right] dt, \tag{5}$$

i.e., we can estimate the probability that $\mathcal{N}(H_d(n,p))$ deviates from some value ν by at most $\gamma\sigma$. However, it is impossible to derive from (4) or (5) the asymptotic probability that $\mathcal{N}(H_d(n,p))$ *hits ν exactly*.

By contrast, our next theorem shows that for any integer ν such that $|\nu - (1-\rho)n| \le O(\sigma)$ we have

$$\mathrm{P}\left[\mathcal{N}(H_d(n,p)) = \nu\right] \sim \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\nu - (1-\rho)n)^2}{2\sigma^2}\right], \tag{6}$$

provided that $(d-1)^{-1} + \varepsilon \le \binom{n-1}{d-1}p = O(1)$. Note that (6) is exactly what we would obtain from (5) if we were allowed to set $\delta = \frac{1}{2}\sigma(n,p)^{-1}$ in that equation. Stated rigorously, the local limit theorem reads as follows.

Theorem 2. *Let $d \ge 2$ be a fixed integer. For any two compact intervals $\mathcal{I} \subset \mathbf{R}$, $\mathcal{J} \subset ((d-1)^{-1}, \infty)$, and for any $\delta > 0$ there exists $n_0 > 0$ such that the following holds. Let $p = p(n)$ be a sequence such that $c = c(n) = \binom{n-1}{d-1}p \in \mathcal{J}$ for all n, let $0 < \rho = \rho(n) < 1$ be the unique solution to (2), and let σ be as in (3). If $n \ge n_0$ and if ν is an integer such that $\sigma^{-1}(\nu - (1-\rho)n) \in \mathcal{I}$, then*

$$\frac{1-\delta}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\nu - (1-\rho)n)^2}{2\sigma^2}\right] \le \mathrm{P}\left[\mathcal{N}(H_d(n,p)) = \nu\right] \le \frac{1+\delta}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\nu - (1-\rho)n)^2}{2\sigma^2}\right].$$

Related work. Since the work of Erdős and Rényi [8], the component structure of $G(n,p) = H_2(n,p)$ has received considerable attention. Stepanov [19] provided central and local limit theorems for $\mathcal{N}(G(n,p))$, thereby proving the $d = 2$ case of Theorems 1 and 2. In order to establish these limit theorems, he estimates the probability that a random graph $G(n,p)$ is connected up to a factor $1 + o(1)$ using recurrence formulas for the number of connected graphs. Furthermore, Barraez, Boucheron, and Fernandez de la Vega [2] reproved the central limit theorem for $\mathcal{N}(G(n,p))$ via the analogy of breadth first search on a random graph and a Galton-Watson branching process. In addition, a local limit theorem for $\mathcal{N}(G(n,p))$ can also be derived using the techniques of van der Hofstad and Spencer [9], or the enumerative results of either Bender, Canfield, and McKay [5] or Pittel and Wormald [15].

Moreover, Pittel [14] proved a central limit theorem for the largest component in the $G(n,m)$ model of random graphs; $G(n,m)$ is just a uniformly distributed graph with exactly n vertices and m edges. Indeed, Pittel actually obtained his central limit theorem via a limit theorem for the joint distribution of the number of isolated trees of a given order, cf. also Janson [10]. A comprehensive treatment of further results on the components of $G(n,p)$ can be found in [11].

In contrast to the case of graphs, only little is known for d-uniform hypergraphs with $d > 2$; for the methods used for graphs do not extend to hypergraphs directly. Using the result [12] on the number of sparsely connected hypergraphs, Karoński and Łuczak [13] investigated the phase transition of $H_d(n,p)$. They established (among other things) a local limit theorem for $\mathcal{N}(H_d(n,m))$ for $m = n/d(d-1) + l$ and $1 \ll \frac{l^3}{n^2} \leq \frac{\ln n}{\ln \ln n}$ which is similar to $H_d(n,p)$ at the regime $\binom{n-1}{d-1}p = (d-1)^{-1} + \omega$, where $n^{-1/3} \ll \omega = \omega(n) \ll n^{-1/3} \ln n / \ln \ln n$. These results were extended by Andriamampianina, Ravelomanana and Rijamamy [1, 16] to the regime $l = o(n^{1/3})$ ($\omega = o(n^{-2/3})$ respectively).

By comparison, Theorems 1 and 2 deal with edge probabilities p such that $\binom{n-1}{d-1}p = (d-1)^{-1} + \Omega(1)$, i.e., $\binom{n-1}{d-1}p$ is bounded away from the critical point $(d-1)^{-1}$. Thus, Theorems 1 and 2 complement [1, 13, 16]. The only prior paper dealing with $\binom{n-1}{d-1}p = (d-1)^{-1} + \Omega(1)$ is that of Coja-Oghlan, Moore, and Sanwalani [7], where the authors computed the expectation and the variance of $\mathcal{N}(H_d(n,p))$ and obtained qualitative results on the component structure of $H_d(n,p)$. In addition, in [7] the authors estimated the probability that $H_d(n,p)$ or a uniformly distributed d-uniform hypergraph $H_d(n,m)$ with n vertices and m edges is connected up to a constant factor. While in the present work we build upon the results on the component structure of $H_d(n,p)$ from [7], the results and techniques of [7] by themselves are not strong enough to obtain a central or even a local limit theorem for $\mathcal{N}(H_d(n,p))$.

Techniques and outline. The aforementioned prior work [1, 12, 13] on the giant component for random hypergraphs relies on enumerative techniques to a significant extent; for the basis [1, 12, 13] are results on the asymptotic number of connected hypergraphs with a given number of vertices and edges. By contrast, in the present work we employ neither enumerative techniques nor results, but rely solely on probabilistic methods. Our proof methods are also quite different from Stepanov's [19], who first estimates the asymptotic probability that a random graph $G(n,p)$ is connected in order to determine the distribution of $\mathcal{N}(H_d(n,p))$. By contrast, in the present work we prove the local limit theorem for $\mathcal{N}(H_d(n,p))$ directly, thereby obtaining "en passant" a new proof for the local limit theorem for random graphs $G(n,p)$, which may be of independent interest. Besides, the local limit theorem can be used to compute the asymptotic probability that $G(n,p)$ or, more generally, $H_d(n,p)$ is connected, or to compute the asymptotic number of connected hypergraphs with a given number of vertices and edges (cf. Section 6). Hence, the general approach taken in the present work is actually converse to the prior ones [1, 12, 13, 19].

The proof of Theorem 1 makes use of *Stein's method*, which is a general technique for proving central limit theorems [18]. Roughly speaking, Stein's result implies that a sum of a family of dependent random variables converges to the normal distribution if one can bound the correlations within any constant-sized subfamily sufficiently well. The method was used by Barbour, Karoński, and Ruciński [3] in order to prove that in a random graph $G(n,p)$, e.g., the number of tree components of a given (bounded) size is asymptotically normal. To establish Theorem 1, we extend their techniques in two ways.

– Instead of dealing with the number of vertices in trees of a given size, we apply Stein's method to the *total* number $n - \mathcal{N}(H_d(n,p))$ of vertices outside of the giant component; this essentially means that we need to sum over all possible (hyper)tree sizes up to about $\ln n$.

– Since we are dealing with hypergraphs rather than graphs, we are facing a somewhat more complex situation than [3], because the fact that an edge may involve an arbitrary number d of vertices yields additional dependencies.

The main contribution of this paper is the proof of Theorem 2. To establish this result, we think of the edges of $H_d(n,p)$ as being added in two “portions”. More precisely, we first include each possible edge with probability $p_1 = (1-\varepsilon)p$ independently, where $\varepsilon > 0$ is small but independent of n (and denote the resulting random hypergraph by H_1); by Theorem 1, the order $\mathcal{N}(H_1)$ of the largest component of H_1 is asymptotically normal. Then, we add each possible edge that is not present in H_1 with a small probability $p_2 \sim \varepsilon p$ and investigate closely how these additional random edges attach further vertices to the largest component of H_1. Denoting the number of these “attached” vertices by $\mathcal{S}$, we will show that the conditional distribution of $\mathcal{S}$ *given the value of* $\mathcal{N}(H_1)$ satisfies a local limit theorem. Since p_1 and p_2 are chosen such that each edge is present with probability p after the second portion of edges has been added, this yields the desired result on $\mathcal{N}(H_d(n,p))$.

The analysis of the conditional distribution of $\mathcal{S}$ involves proving that $\mathcal{S}$ is asymptotically normal. To show this, we employ Stein’s method once more. In addition, in order to show that $\mathcal{S}$ satisfies a *local* limit theorem, we prove that the number of isolated vertices of H_1 that get attached to the largest component of H_1 by the second portion of random edges is binomially distributed. Since the binomial distribution satisfies a local limit theorem, we thus obtain a local limit theorem for $\mathcal{S}$.

Our proof of Theorem 2 makes use of some results on the component structure of $H_d(n,p)$ derived in [7]. For instance, we employ the results on the expectation and the variance of $\mathcal{N}(H_d(n,p))$ from that paper. Furthermore, the analysis of $\mathcal{S}$ given in the present work is a considerable extension of the argument used in [7], which by itself would just yield the probability that $\mathcal{S}$ attains a specific value s *up to a constant factor*.

The main part of the paper is organized as follows. After making some preliminaries in Section 2, we outline the proof of Theorem 2 in Section 3. In that section we explain in detail how $H_d(n,p)$ is generated in two “portions”. Then, in Section 4 we analyze the random variable $\mathcal{S}$, assuming the central limit theorem for $\mathcal{S}$. Further, Section 5 deals with the proof of Theorem 1 and the proof of the central limit theorem for $\mathcal{S}$ via Stein’s method; the reason why we defer the proof of Theorem 1 to Section 5 is that we can use basically the same argument to prove the asymptotic normality of both $\mathcal{N}(H_d(n,p))$ and $\mathcal{S}$. Finally, Section 6 contains some concluding remarks, e.g., on the use of the present results to derive further limit theorems and to solve enumerative problems.

2 Preliminaries

Throughout the paper, we let $V = \{1,\ldots,n\}$. If $d \geq 2$ is an integer and $V_1,\ldots,V_k \subset V$, then we let $\mathcal{E}_d(V_1,\ldots,V_k)$ signify the set of all subsets $e \subset V$ of cardinality d such that $e \cap V_i \neq \emptyset$ for all i. We omit the subscript d if it is clear from the context.

If H is a hypergraph, then we let $V(H)$ denote its vertex set and $E(H)$ its edge set. We say that a set $S \subset V(H)$ is *reachable from* $T \subset V(H)$ if each vertex $s \in S$ is reachable from some vertex $t \in T$. Further, if $V(H) \subset V = \{1,\ldots,n\}$, then the subsets of V can be ordered lexicographically; hence, we can define the *largest component* of H to be the lexicographically first component of order $\mathcal{N}(H)$.

We use the O-notation to express asymptotic estimates as $n \to \infty$. Furthermore, if $f(x_1,\ldots,x_k,n)$ is a function that depends not only on n but also on some further parameters x_i from domains $D_i \subset \mathbf{R}$ ($1 \leq i \leq k$), and if $g(n) \geq 0$ is another function, then we say that the estimate $f(x_1,\ldots,x_k,n) = O(g(n))$ holds *uniformly in* $x_1,\ldots,x_k$ if the following is true: if $\mathcal{I}_j$ and D_j, $\mathcal{I}_j \subset D_j$, are compact sets, then there exist numbers $C = C(\mathcal{I}_1,\ldots,\mathcal{I}_k)$ and $n_0 = n_0(\mathcal{I}_1,\ldots,\mathcal{I}_k)$ such that $|f(x_1,\ldots,x_k,n)| \leq Cg(n)$ for all $n \geq n_0$ and $(x_1,\ldots,x_k) \in \prod_{j=1}^k \mathcal{I}_j$. We define uniformity analogously for the other Landau symbols Ω, Θ, etc.

We shall make repeated use of the following *Chernoff bound* on the tails of a binomially distributed variable $X = \mathrm{Bin}(\nu,q)$ (cf. [11, p. 26] for a proof): for any $t > 0$ we have

$$\mathrm{P}\left[|X - \mathrm{E}(X)| \geq t\right] \leq 2\exp\left(-\frac{t^2}{2(\mathrm{E}(X) + t/3)}\right). \tag{7}$$

Moreover, we employ the following *local limit theorem* for the binomial distribution (cf. [6, Chapter 1]).

Proposition 3. *Suppose that $0 \le p = p(n) \le 1$ is a sequence such that $np(1-p) \to \infty$ as $n \to \infty$. Let $X = \mathrm{Bin}(n,p)$. Then for any sequence $x = x(n)$ of integers such that $|x - np| = o(np(1-p))^{2/3}$,*

$$\mathrm{P}\left[X = x\right] \sim (2\pi np(1-p))^{-\frac{1}{2}} \exp\left(-\frac{(x-np)^2}{2p(1-p)n}\right) \qquad \text{as } n \to \infty.$$

Furthermore, we make use of the following theorem, which summarizes results from [7, Section 6] on the component structure of $H_d(n,p)$.

Theorem 4. *Let $p = c\binom{n-1}{d-1}^{-1}$.*

1. *If there is a fixed $c_0 < (d-1)^{-1}$ such that $c = c(n) \le c_0$, then*

$$\mathrm{P}\left[\mathcal{N}(H_d(n,p)) \le 3(d-1)^2(1-(d-1)c_0)^{-2}\ln n\right] \ge 1 - n^{-100}.$$

2. *Suppose that $c_0 > (d-1)^{-1}$ is a constant, and that $c_0 \le c = c(n) = o(\ln n)$ as $n \to \infty$. Then the transcendental equation (2) has a unique solution $0 < \rho = \rho(c) < 1$, which satisfies*

$$\binom{\rho n}{d-1} p < c_0' < (d-1)^{-1}. \tag{8}$$

for some number $c_0' > 0$ that depends only on c_0. Moreover,

$$\begin{aligned} \left|\mathrm{E}\left[\mathcal{N}(H_d(n,p))\right] - (1-\rho)n\right| &\le n^{o(1)}, \\ \mathrm{Var}(\mathcal{N}(H_d(n,p))) &\sim \frac{\rho\left[1-\rho+c(d-1)(\rho-\rho^{d-1})\right]n}{(1-c(d-1)\rho^{d-1})^2}. \end{aligned}$$

Furthermore, with probability $\ge 1-n^{-100}$ there is precisely one component of order $(1+o(1))(1-\rho)n$ in $H_d(n,p)$, while all other components have order $\le \ln^2 n$. In addition,

$$\mathrm{P}\left[\left|\mathcal{N}(H_d(n,p)) - \mathrm{E}(\mathcal{N}(H_d(n,p)))\right| \ge n^{0.51}\right] \le n^{-100}.$$

Finally, the following result on the component structure of $H_d(n,p)$ with average degree $\binom{n-1}{d-1}p < (d-1)^{-1}$ below the threshold has been derived in [7, Section 6] via the theory of branching processes.

Proposition 5. *There exists a function $q : (0,(d-1)^{-1}) \times [0,1] \to \mathbf{R}_{\ge 0}$, $(\zeta,\xi) \mapsto q(\zeta,\xi) = \sum_{k=1}^{\infty} q_k(\zeta)\xi^k$ whose coefficients $\zeta \mapsto q_k(\zeta)$ are differentiable such that the following holds. Suppose that $0 \le p = p(n) \le 1$ is a sequence such that $0 < \binom{n-1}{d-1}p = c = c(n) < (d-1)^{-1} - \varepsilon$ for an arbitrarily small $\varepsilon > 0$ that remains fixed as $n \to \infty$. Let $P(c,k)$ denote the probability that in $H_d(n,p)$ some fixed vertex $v \in V$ lies in a component of order k. Then*

$$P(c,k) = (1+o(n^{-2/3}))q_k(c) \qquad \text{for all } 1 \le k \le \ln^2 n.$$

Furthermore, for any fixed $\varepsilon > 0$ there is a number $0 < \gamma = \gamma(\varepsilon) < 1$ such that

$$q_k(c) \le \gamma^k \quad \text{for all } 0 < c < (d-1)^{-1} - \varepsilon. \tag{9}$$

3 Proof of Theorem 2

Throughout this section, we assume that $c = c(n) = \binom{n-1}{d-1}p \in \mathcal{J}$ for some compact interval $\mathcal{J} \subset ((d-1)^{-1}, \infty)$. Moreover, we let $\mathcal{I} \subset \mathbf{R}$ be some fixed compact interval, and ν denotes an integer such that $(\nu - (1-\rho)n)/\sigma \in \mathcal{I}$. All asymptotics are understood to hold uniformly in c and $(\nu - (1-\rho)n)/\sigma$.

3.1 Outline

Let $\varepsilon = \varepsilon(\mathcal{J}) > 0$ be independent of n and small enough so that $(1-\varepsilon)\binom{n-1}{d-1}p > (d-1)^{-1} + \varepsilon$. Set $p_1 = (1-\varepsilon)p$. Moreover, let p_2 be the solution to the equation $p_1 + p_2 - p_1p_2 = p$; then $p_2 \sim \varepsilon p$. We expose the edges of $H_d(n,p)$ in four "rounds" as follows.

R1. As a first step, we let H_1 be a random hypergraph obtained by including each of the $\binom{n}{d}$ possible edges with probability p_1 independently. Let G denote the largest component of H_1.
R2. Let H_2 be the hypergraph obtained from H_1 by adding each edge $e \notin H_1$ that lies completely outside of G (i.e., $e \subset V \setminus G$) with probability p_2 independently.
R3. Obtain H_3 by adding each possible edge $e \notin H_1$ that contains vertices of both G and $V \setminus G$ with probability p_2 independently.
R4. Finally, include each possible edge $e \notin H_1$ such that $e \subset G$ with probability p_2 independently.

Here the 1st round corresponds to the first portion of edges mentioned in Section 1, and the edges added in the 2nd–4th round correspond to the second portion. Note that for each possible edge $e \subset V$ the probability that e is actually present in H_4 is $p_1 + (1-p_1)p_2 = p$, hence $H_4 = H_d(n,p)$. Moreover, as $\binom{n-1}{d-1}p_1 > (d-1)^{-1} + \varepsilon$ by our choice of ε, Theorem 4 entails that w.h.p. H_1 has exactly one largest component of linear size $\Omega(n)$ (the "giant component"). Further, the edges added in the 4th round do not affect the order of the largest component, i.e., $\mathcal{N}(H_4) = \mathcal{N}(H_3)$.

In order to analyze the distribution of $\mathcal{N}(H_d(n,p))$, we first establish *central limit theorems* for $\mathcal{N}(H_1) = |G|$ and $\mathcal{N}(H_3) = \mathcal{N}(H_4) = \mathcal{N}(H_d(n,p))$, i.e., we prove that (centralized and normalized versions of) $\mathcal{N}(H_1)$ and $\mathcal{N}(H_3)$ are asymptotically normal. Then, we investigate the number of vertices $\mathcal{S} = \mathcal{N}(H_3) - \mathcal{N}(H_1)$ that get attached to G_1 during the 3rd round. We shall prove that *given that* $|G| = n_1$, $\mathcal{S}$ is *locally* normal with mean $\mu_{\mathcal{S}} + (n_1 - \mu_1)\lambda_{\mathcal{S}}$ and variance $\sigma_{\mathcal{S}}^2$ independent of n_1. Finally, we combine these results to obtain the local limit theorem for $\mathcal{N}(H_d(n,p)) = \mathcal{N}(H_3) = \mathcal{N}(H_1) + \mathcal{S}$.

Let $c_1 = \binom{n-1}{d-1}p_1$ and $c_3 = \binom{n-1}{d-1}p$. Moreover, let $0 < \rho_3 < \rho_1 < 1$ signify the solutions to the transcendental equations $\rho_j = \exp\left[c_j(\rho_j^{d-1} - 1)\right]$ and set for $j = 1, 3$

$$\mu_j = (1-\rho_j)n, \quad \sigma_j^2 = \frac{\rho_j\left[1 - \rho_j + c_j(d-1)(\rho_j - \rho_j^{d-1})\right]n}{(1 - c_j(d-1)\rho_j^{d-1})^2} \quad \text{(cf. Theorem 4).}$$

The following proposition, which we will prove in Section 5, establishes a central limit theorem for both $\mathcal{N}(H_1)$ and $\mathcal{N}(H_3)$ and thus proves Theorem 1.

Proposition 6. *$(\mathcal{N}(H_j) - \mu_j)/\sigma_j$ converges in distribution to the standard normal distribution for $j = 1, 3$.*

With respect to the distribution of $\mathcal{S}$, we will establish the following local limit theorem in Section 4.

Proposition 7. *Suppose that $|n_1 - \mu_1| \le n^{0.6}$.*

1. *The conditional expectation of $\mathcal{S}$ given that $|G| = n_1$ satisfies $\mathrm{E}(\mathcal{S}|\mathcal{N}_1 = n_1) = \mu_{\mathcal{S}} + \lambda_{\mathcal{S}}(n_1 - \mu_1) + o(\sqrt{n})$, where $\mu_{\mathcal{S}} = \Theta(n)$ and $\lambda_{\mathcal{S}} = \Theta(1)$ are independent of n_1.*
2. *There is a constant $C > 0$ such that for all s satisfying $|\mu_{\mathcal{S}} + \lambda_{\mathcal{S}}(n_1 - \mu_1) - s| \le n^{0.6}$ we have $\mathrm{P}\left[\mathcal{S} = \nu|\mathcal{N}_1 = n_1\right] \le Cn^{-\frac{1}{2}}$.*
3. *If s is an integer such that $|\mu_{\mathcal{S}} + \lambda_{\mathcal{S}}(n_1 - \mu_1) - s| \le O(\sqrt{n})$, then*

$$\mathrm{P}\left[\mathcal{S} = s|\mathcal{N}_1 = n_1\right] \sim \frac{1}{\sqrt{2\pi}\sigma_{\mathcal{S}}} \exp\left(-\frac{(\mu_{\mathcal{S}} + \lambda_{\mathcal{S}}(n_1 - \mu_1) - s)^2}{2\sigma_{\mathcal{S}}^2}\right),$$

where $\sigma_{\mathcal{S}} = \Theta(\sqrt{n})$ is independent of n_1.

Since $\mathcal{N}_3 = \mathcal{N}_1 + \mathcal{S}$, Propositions 6 and 7 yield

$$\mu_3 = \mu_1 + \mu_{\mathcal{S}} + o(\sqrt{n}). \tag{10}$$

Combining Propositions 6 and 7, we derive the following formula for $\mathrm{P}\left[\mathcal{N}_3 = \nu\right]$ in Section 3.2. Recall that we are assuming that ν is an integer such that $(\nu - \mu)/\sigma = (\nu - \mu_3)/\sigma_3 \in \mathcal{I}$.

Corollary 8. *Letting $z = (\nu - \mu_3)/\sigma_3$, we have*

$$\mathrm{P}\left[\mathcal{N}_3 = \nu\right] \sim \frac{1}{2\pi\sigma_{\mathcal{S}}} \int_{-\infty}^{\infty} \exp\left[-\frac{x^2}{2} - \frac{1}{2}\left((x \cdot (1+\lambda_{\mathcal{S}})\frac{\sigma_1}{\sigma_{\mathcal{S}}} - z \cdot \frac{\sigma_3}{\sigma_{\mathcal{S}}}\right)^2\right] dx. \tag{11}$$

Proof of Theorem 2. Integrating the right hand side of (11), we obtain an expression of the form

$$\mathrm{P}\left[\mathcal{N}_3 = \nu\right] \sim \frac{1}{\sqrt{2\pi}\tau} \exp\left(-\frac{(\nu-\kappa)^2}{2\tau^2}\right), \tag{12}$$

where $\kappa, \tau^2 = \Theta(n)$. Therefore, on the one hand $(\mathcal{N}_3 - \mu_3)/\sigma_3$ converges in distribution to the normal distribution with mean $\kappa - \mu_3$ and variance $(\tau/\sigma_3)^2$. On the other hand, Proposition 6 states that $(\mathcal{N}_3 - \mu_3)/\sigma_3$ converges to the standard normal distribution. Consequently, $|\kappa - \mu_3| = o(\tau)$ and $\tau \sim \sigma_3$. Plugging these estimates into (12), we obtain $\mathrm{P}\left[\mathcal{N}_3 = \nu\right] \sim \frac{1}{\sqrt{2\pi}\sigma_3} \exp\left(-\frac{1}{2}(\nu - \mu_3)^2 \sigma_3^{-2}\right)$. Since $\mathcal{N}_3 = \mathcal{N}(H_d(n,p))$, this yields the assertion. □

3.2 Proof of Corollary 8

Let $\alpha > 0$ be arbitrarily small but fixed as $n \to \infty$, and let $C' = C'(\alpha) > 0$ be a large enough number depending only on α. Set $J = \{n_1 \in \mathbb{Z} : |n_1 - \mu_1| \le C'\sqrt{n}\}$, let $J' = \{n_1 \in \mathbb{Z} : C'\sqrt{n} < |n_1 - \mu_1| \le n^{0.6}\}$, and $J'' = \{n_1 \in \mathbb{Z} : |n_1 - \mu_1| > n^{0.6}\}$. Then letting

$$\Psi_X = \sum_{n_1 \in X} \mathrm{P}\left[\mathcal{N}_1 = n_1\right] \mathrm{P}\left[\mathcal{S} = \nu - n_1 | \mathcal{N}_1 = n_1\right], \quad \text{for } X \in \{J, J', J''\}$$

we have $\mathrm{P}\left[\mathcal{N}_3 = \nu\right] = \Psi_J + \Psi_{J'} + \Psi_{J''}$, and we shall estimate each of the three summands individually.

Since Theorem 4 implies that $\mathrm{P}\left[|\mathcal{N}_1 - \mu_1| > n^{0.51}\right] \le n^{-100}$, we conclude that

$$\Psi_{J''} \le \mathrm{P}\left[\mathcal{N}_1 \in J''\right] \le n^{-100}. \tag{13}$$

Furthermore, as $\sigma_1^2 = O(n)$, Chebyshev's inequality implies that

$$\mathrm{P}\left[\mathcal{N}_1 \in J'\right] \le \mathrm{P}\left[|\mathcal{N}_1 - \mu_1| > C'\sqrt{n}\right] \le \sigma_1^2 C'^{-2} n^{-1} < \alpha/C', \tag{14}$$

provided that C' is large enough. Hence, combining (14) with the second part of Proposition 7, we obtain

$$\Psi_{J'} \le \mathrm{P}\left[\mathcal{N}_1 \in J'\right] \cdot \frac{C}{\sqrt{n}} \le \frac{\alpha C}{C'\sqrt{n}} < \alpha n^{-1/2}, \tag{15}$$

where once we need to pick C' sufficiently large.

To estimate the contribution of $n_1 \in J$, we split J into subintervals $J_1, \ldots, J_K$ of length between $\frac{\sigma_1}{2C'}$ and $\frac{\sigma_1}{C'}$. Moreover, let I_j be the interval $[(\min J_j - \mu_1)/\sigma_1, (\max J_j - \mu_1)/\sigma_1]$. Then Proposition 6 implies that

$$\frac{1-\alpha}{\sqrt{2\pi}} \int_{I_j} \exp(-x^2/2) dx \le \sum_{n_1 \in J_j} \mathrm{P}\left[\mathcal{N}_1 = n_1\right] \le \frac{1+\alpha}{\sqrt{2\pi}} \int_{I_j} \exp(-x^2/2) dx \tag{16}$$

for each $1 \le j \le K$. Furthermore, Proposition 7 yields

$$\mathrm{P}\left[\mathcal{S} = \nu - n_1 | \mathcal{N}_1 = n_1\right] \sim \frac{1}{\sqrt{2\pi}\sigma_{\mathcal{S}}} \exp\left(-\frac{(\nu - n_1 - \mu_{\mathcal{S}} - \lambda_{\mathcal{S}}(n_1 - \mu_1))^2}{2\sigma_{\mathcal{S}}^2}\right).$$

for each $n_1 \in J$. Hence, choosing C' sufficiently large, we can achieve that for all $n_1 \in J_j$ and all $x \in I_j$ the bound

$$\begin{aligned}\mathrm{P}\left[\mathcal{S} = \nu - n_1 | \mathcal{N}_1 = n_1\right] &\le \frac{(1+\alpha)^2}{\sqrt{2\pi}\sigma_{\mathcal{S}}} \exp\left(-\frac{(\nu - \mu_1 - \sigma_1 x - \mu_{\mathcal{S}} - \lambda_{\mathcal{S}}(n_1 - \mu_1))^2}{2\sigma_{\mathcal{S}}^2}\right) \\ &\overset{(10)}{\sim} \frac{(1+\alpha)^2}{\sqrt{2\pi}\sigma_{\mathcal{S}}} \exp\left(-\frac{1}{2}\left((x \cdot (1+\lambda_{\mathcal{S}})\frac{\sigma_1}{\sigma_{\mathcal{S}}} - z \cdot \frac{\sigma_3}{\sigma_{\mathcal{S}}}\right)^2\right)\end{aligned} \tag{17}$$

holds. Now, combining (16) and (17), we conclude that

$$\begin{aligned}\Psi_J &= \sum_{j=1}^{K} \sum_{n_1 \in J_j} \mathrm{P}\left[\mathcal{N}_1 = n_1\right] \mathrm{P}\left[\mathcal{S} = \nu - n_1 | \mathcal{N}_1 = n_1\right] \\ &\leq \frac{(1+\alpha)^3}{2\pi\sigma_{\mathcal{S}}} \sum_{j=1}^{K} \int_{I_j} \exp\left[-\frac{x^2}{2} - \frac{1}{2}\left((x \cdot (1+\lambda_{\mathcal{S}})\frac{\sigma_1}{\sigma_{\mathcal{S}}} - z \cdot \frac{\sigma_3}{\sigma_{\mathcal{S}}}\right)^2\right] dx \\ &\leq \frac{1+4\alpha}{2\pi\sigma_{\mathcal{S}}} \int_{-\infty}^{\infty} \exp\left[-\frac{x^2}{2} - \frac{1}{2}\left((x \cdot (1+\lambda_{\mathcal{S}})\frac{\sigma_1}{\sigma_{\mathcal{S}}} - z \cdot \frac{\sigma_3}{\sigma_{\mathcal{S}}}\right)^2\right] dx. \end{aligned} \tag{18}$$

Analogously, we derive the matching lower bound

$$\Psi_J \geq \frac{1-4\alpha}{2\pi\sigma_{\mathcal{S}}} \int_{-\infty}^{\infty} \exp\left[-\frac{x^2}{2} - \frac{1}{2}\left((x \cdot (1+\lambda_{\mathcal{S}})\frac{\sigma_1}{\sigma_{\mathcal{S}}} - z \cdot \frac{\sigma_3}{\sigma_{\mathcal{S}}}\right)^2\right] dx. \tag{19}$$

Finally, combining (13), (15), (18), and (19), and remembering that $\mathrm{P}\left[\mathcal{N}_3 = \nu\right] = \Psi_J + \Psi_{J'} + \Psi_{J''}$, we obtain the assertion, because $\alpha > 0$ can be chosen arbitrarily small if n gets sufficiently large.

4 The Conditional Distribution of $\mathcal{S}$

Throughout this section, we keep the notation and the assumptions from Section 3. In addition, we let $G \subset V$ be a set of cardinality n_1 such that $|n_1 - \mu_1| \leq n^{0.6}$.

4.1 Outline

The goal of this section is to prove Proposition 7. Let us condition on the event that the largest component of H_1 is G. To analyze the conditional distribution of $\mathcal{S}$, we need to overcome the problem that in H_1 the edges in the set $V \setminus G$ do not occur independently anymore once we condition on G being the largest component of H_1. However, we will see that this conditioning is “not very strong”. To this end, we shall compare $\mathcal{S}$ with an “artificial” random variable $\mathcal{S}_G$, which models the edges contained in $V \setminus G$ as mutually independent objects. To define $\mathcal{S}_G$, we set up random hypergraphs $H_{j,G}$, $j = 1, 2, 3$, in three “rounds” as follows.

R1’. The vertex set of $H_{1,G}$ is $V = \{1, \ldots, n\}$, and each of the $\binom{n-n_1}{d}$ possible edges $e \subset V \setminus G$ is present in $H_{1,G}$ with probability p_1 independently.
R2’. Adding each possible edge $e \subset V \setminus G$ not present in $H_{1,G}$ with probability p_2 independently yields $H_{2,G}$.
R3’. Obtain $H_{3,G}$ from $H_{2,G}$ by including each possible edge e incident to both G and $V \setminus G$ with probability p_2 independently.

The process R1’–R3’ relates to the process R1–R4 from Section 3.1 as follows. While in H_1 the edges in $V \setminus G$ are mutually dependent, we have “artificially” constructed $H_{1,G}$ in such a way that the edges outside of G occur independently. Then, $H_{2,G}$ and $H_{3,G}$ are obtained similarly as H_2 and H_3, namely by including further edges inside of $V \setminus G$ and crossing edges between G and $V \setminus G$ with probability p_2. Letting S_G denote the set of vertices in $V \setminus G$ that are reachable from G, the quantity $\mathcal{S}_G = |S_G|$ now corresponds to $\mathcal{S}$. In contrast to R1–R4, the process R1’–R3’ completely disregards edges inside of G, because these do not affect $\mathcal{S}_G$. The following lemma, which we will prove in Section 4.3 shows that $\mathcal{S}_G$ is indeed a very good approximation of $\mathcal{S}$, so that it suffices to study $\mathcal{S}_G$.

Lemma 9. *For any $\nu \in \mathbb{Z}$ we have $|\mathrm{P}\left[\mathcal{S} = \nu \mid \mathcal{N}(H_1) = n_1\right] - \mathrm{P}\left[\mathcal{S}_G = \nu\right]| \leq n^{-9}$.*

As a next step, we investigate the expectation of $\mathcal{S}_G$. While there is no need to compute $\mathrm{E}(\mathcal{S}_G)$ precisely, we do need that $\mathrm{E}(\mathcal{S}_G)$ depends on $n_1 - \mu_1$ *linearly*. The corresponding proof can be found in Section 4.4.

Lemma 10. *We have* $\mathrm{E}(\mathcal{S}_G) = \mu_{\mathcal{S}} + \lambda_{\mathcal{S}}(n_1 - \mu_1) + o(\sqrt{n})$, *where* $\mu_{\mathcal{S}} = \Theta(n)$ *and* $\lambda_{\mathcal{S}} = \Theta(1)$ *do not depend on* n_1.

Furthermore, we need that the variance of $\mathcal{S}_G$ is essentially independent of the precise value of n_1. This will be proven in Section 4.5.

Lemma 11. *We have* $\mathrm{Var}(\mathcal{S}_G) = O(n)$. *Moreover, if* $G' \subset V$ *is another set such that* $|\mu_1 - |G'|| = o(n)$, *then* $|\mathrm{Var}(\mathcal{S}_G) - \mathrm{Var}(\mathcal{S}_{G'})| = o(n)$.

To show that $\mathcal{S}_G$ satisfies a local limit theorem, the crucial step is to prove that for numbers s and t such that s is "close" to t the probabilities $\mathrm{P}[\mathcal{S}_G = s]$, $\mathrm{P}[\mathcal{S}_G = t]$ are "almost the same". More precisely, the following lemma, proven in Section 4.2, holds.

Lemma 12. *For every* $\alpha > 0$ *there is* $\beta > 0$ *such that for all* s, t *satisfying* $|s - \mathrm{E}(\mathcal{S}_G)|, |t - \mathrm{E}(\mathcal{S}_G)| \leq n^{0.6}$ *and* $|s - t| \leq \beta n^{1/2}$ *we have*

$$(1-\alpha)\mathrm{P}[\mathcal{S}_G = s] - n^{-10} \leq \mathrm{P}[\mathcal{S}_G = t] \leq (1+\alpha)\mathrm{P}[\mathcal{S}_G = s] + n^{-10}.$$

Moreover, there is a constant $C > 0$ *such that* $\mathrm{P}[\mathcal{S}_G = s] \leq Cn^{-1/2}$ *for all integers* s.

Letting $G_0 = \{1, \ldots, \lceil \mu_1 \rceil\}$, we define $\sigma_{\mathcal{S}}^2 = \mathrm{Var}(\mathcal{S}_{G_0})$ and obtain a lower bound on $\sigma_{\mathcal{S}}$ as an immediate consequence of Lemma 12.

Corollary 13. *We have* $\sigma_{\mathcal{S}} = \Omega(\sqrt{n})$.

Proof. By Lemma 12 there exists a number $0 < \beta < 0.01$ independent of n such that for all integers s, t satisfying $|s - \mathrm{E}(\mathcal{S}_G)|, |t - \mathrm{E}(\mathcal{S}_G)| \leq \sqrt{n}$ and $|s - t| \leq \beta\sqrt{n}$ we have

$$\mathrm{P}[\mathcal{S}_G = t] \geq \frac{2}{3}\mathrm{P}[\mathcal{S}_G = s] - n^{-10}. \tag{20}$$

Set $\gamma = \beta^2/64$ and assume for contradiction that $\sigma_{\mathcal{S}}^2 < \gamma n/2$. Moreover, suppose that $G = G_0 = \{1, \ldots, \lceil \mu_1 \rceil\}$. Then Chebyshev's inequality entails that $\mathrm{P}\left[|\mathcal{S}_G - \mathrm{E}(\mathcal{S}_G)| \geq \sqrt{\gamma n}\right] \leq \frac{1}{2}$. Hence, there exists an integer s such that $|s - \mathrm{E}(\mathcal{S}_G)| \leq \sqrt{\gamma n}$ and $\mathrm{P}[\mathcal{S}_G = s] \geq \frac{1}{2}(\gamma n)^{-\frac{1}{2}}$. Therefore, due to (20) we have $\mathrm{P}[\mathcal{S}_G = t] \geq \frac{1}{4}(\gamma n)^{-\frac{1}{2}}$ for all integers t such that $|s - t| \leq \beta\sqrt{n}$. Thus, recalling that $\gamma = \beta^2/64$, we obtain $1 \geq \mathrm{P}[|\mathcal{S}_G - s| \leq \beta\sqrt{n}] = \sum_{t:|t-s|\leq\beta\sqrt{n}} \mathrm{P}[\mathcal{S}_G = t] \geq \frac{\beta\sqrt{n}}{4\sqrt{\gamma n}} > 1$. This contradiction shows that $\sigma_{\mathcal{S}}^2 \geq \gamma n/2$. □

Using the above estimates of the expectation and the variance of $\mathcal{S}_G$ and invoking Stein's method once more, in Section 5 we will show the following.

Lemma 14. *If* $|n_1 - \mu_1| \leq n^{0.66}$, *then* $(\mathcal{S}_G - \mathrm{E}(\mathcal{S}_G))/\sigma_{\mathcal{S}}$ *is asymptotically normal.*

Proof of Proposition 7. The first part of the proposition follows readily from Lemmas 9 and 10. Moreover, the second assertion follows from Lemma 12. Furthermore, we shall establish below that

$$\mathrm{P}[\mathcal{S}_G = s] \sim \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(s - \mathrm{E}(\mathcal{S}))^2}{2\sigma_{\mathcal{S}}^2}\right) \quad \text{for any integer } s \text{ such that } |s - \mathrm{E}(\mathcal{S}_G)| = O(\sqrt{n}). \tag{21}$$

This claim implies the third part of the proposition. For $(s - \mathrm{E}(\mathcal{S}))^2\sigma_{\mathcal{S}}^{-2} \sim (\mu_{\mathcal{S}} + \lambda_{\mathcal{S}}(n_1 - \mu_1))^2\sigma_{\mathcal{S}}^{-2}$ by Lemma 10 and Corollary 13, and $\mathrm{P}[\mathcal{S} = s | \mathcal{N}_1 = n_1] \sim \mathrm{P}[\mathcal{S}_G = s]$ by Lemma 9.

To prove (21) let $\alpha > 0$ be arbitrarily small but fixed. Since $\sigma_{\mathcal{S}}^2 = \Theta(n)$ by Lemma 11 and Corollary 13, Lemma 12 entails that for a sufficiently small $\beta > 0$ and all s, t satisfying $|s - \mathrm{E}(\mathcal{S}_G)|, |t - \mathrm{E}(\mathcal{S}_G)| \leq n^{0.6}$ and $|s - t| \leq \beta\sigma_{\mathcal{S}}$ we have

$$(1-\alpha)\mathrm{P}[\mathcal{S}_G = s] - n^{-10} \leq \mathrm{P}[\mathcal{S}_G = t] \leq (1+\alpha)\mathrm{P}[\mathcal{S}_G = s] + n^{-10}. \tag{22}$$

Now, suppose that s is an integer such that $|s - \mathrm{E}(\mathcal{S}_G)| \le O(\sqrt{n})$, and set $z = (s - \mathrm{E}(\mathcal{S}_G))/\sigma_{\mathcal{S}}$. Then Lemma 14 implies that

$$\mathrm{P}\left[|\mathcal{S}_G - s| \le \beta\sigma_{\mathcal{S}}\right] \ge \frac{1-\alpha}{\sqrt{2\pi}} \int_{z-\beta}^{z+\beta} \exp(-x^2/2)dx \ge (1-2\alpha)\frac{\beta}{\sqrt{2\pi}} \exp(-z^2/2), \tag{23}$$

provided that β is small enough. Furthermore, (22) yields that

$$\begin{aligned}\mathrm{P}\left[|\mathcal{S}_G - s| \le \beta\sigma_{\mathcal{S}}\right] &= \sum_{t:|t-s|\le\beta\sigma_{\mathcal{S}}} \mathrm{P}\left[\mathcal{S}_G = t\right] \le \beta\sigma_{\mathcal{S}}((1+\alpha)\mathrm{P}\left[\mathcal{S}_G = s\right] + n^{-10}) \\ &\le (1+\alpha)\beta\sigma_{\mathcal{S}}\mathrm{P}\left[\mathcal{S}_G = s\right] + n^{-9},\end{aligned} \tag{24}$$

because $\sigma_{\mathcal{S}} = O(\sqrt{n})$ by Lemma 11. Combining (23) and (24), we conclude that

$$\mathrm{P}\left[\mathcal{S}_G = s\right] \ge \frac{1-2\alpha}{1+\alpha} \cdot \frac{1}{\sqrt{2\pi}\sigma_{\mathcal{S}}} \exp(-z^2/2) - n^{-9} \ge \frac{1-4\alpha}{\sqrt{2\pi}\sigma_{\mathcal{S}}} \exp\left(-\frac{(s - \mathrm{E}(\mathcal{S}_G))^2}{2\sigma_{\mathcal{S}}^2}\right).$$

Since analogous arguments yield the matching upper bound $\mathrm{P}\left[\mathcal{S}_G = s\right] \le \frac{1+4\alpha}{\sqrt{2\pi}\sigma_{\mathcal{S}}} \exp\left(-\frac{(s-\mathrm{E}(\mathcal{S}_G))^2}{2\sigma_{\mathcal{S}}^2}\right)$, and because $\alpha > 0$ may be chosen arbitrarily small, we obtain (21). $\square$

Next we will prove Lemma 12 which provides the central locality argument while the more technical proofs of Lemma 9, 10 and 11 are deferred to the end of this section.

4.2 Proof of Lemma 12

Since the assertion is symmetric in s and t, it suffices to prove that $\mathrm{P}\left[\mathcal{S}_G = s\right] \le (1-\alpha)^{-1}\mathrm{P}\left[\mathcal{S}_G = s\right] + n^{-10}$. Let $\mathcal{F} = E(H_{3,G}) \setminus E(H_{2,G})$ be the (random) set of edges added during **R3'**. We split $\mathcal{F}$ into three subsets: let $\mathcal{F}_1$ consist of all $e \in \mathcal{F}$ such that either $|e \setminus G| \ge 2$ or e contains a vertex that belongs to a component of $V \setminus G$ of order ≥ 2. Moreover, $\mathcal{F}_2$ is the set of all edges $e \in \mathcal{F} \setminus \mathcal{F}_1$ that contain a vertex of $V \setminus G$ that is also contained in some other edge $e' \in \mathcal{F}_1$. Finally, $\mathcal{F}_3 = \mathcal{F} \setminus (\mathcal{F}_1 \cup \mathcal{F}_2)$; thus, all edges $e \in \mathcal{F}_3$ connect $d-1$ vertices in G with a vertex $v \in V \setminus G$ that is isolated in $H_{2,G} + \mathcal{F}_1 + \mathcal{F}_2$, see Figure 1 for an example. Hence, $H_{3,G} = H_{2,G} + \mathcal{F}_1 + \mathcal{F}_2 + \mathcal{F}_3$.

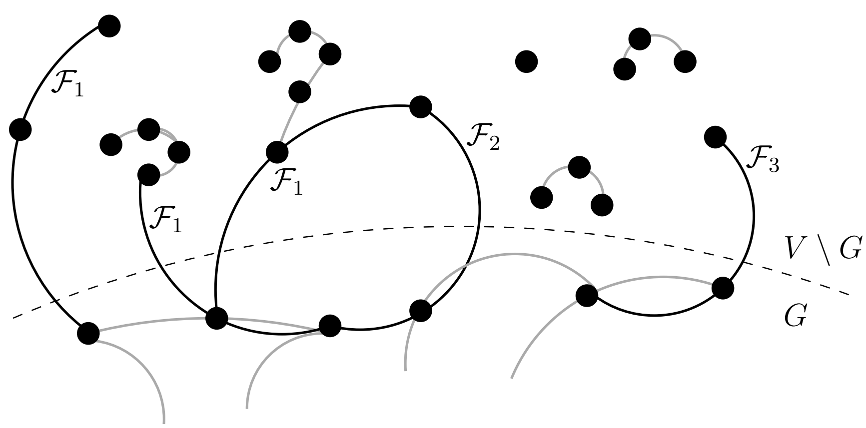

Fig. 1. The three kinds of edges (black) which attach small components to G. The edges of $H_{2,G}$ are depicted in grey. The (3-uniform) edges are depicted as circular arcs spanned by the three vertices contained in the corresponding edge.

As a next step, we decompose $\mathcal{S}_G$ into two contributions corresponding to $\mathcal{F}_1 \cup \mathcal{F}_2$ and $\mathcal{F}_3$. More precisely, we let $\mathcal{S}_G^{\mathrm{big}}$ be the number of vertices in $V \setminus G$ that are reachable from G in $H_{2,G} + \mathcal{F}_1 + \mathcal{F}_2$ and

set $\mathcal{S}_G^{\mathrm{iso}} = \mathcal{S}_G - \mathcal{S}_G^{\mathrm{big}}$. Hence, if we let $\mathcal{W}$ signify the set of all isolated vertices of $H_{2,G} + \mathcal{F}_1 + \mathcal{F}_2$ in the set $V \setminus G$, then $\mathcal{S}_G^{\mathrm{iso}}$ equals the number of vertices in $\mathcal{W}$ that get attached to G via the edges in $\mathcal{F}_3$.

We can determine the distribution of $\mathcal{S}_G^{\mathrm{iso}}$ precisely. For if $v \in \mathcal{W}$, then each edge e containing v and exactly $d-1$ vertices of G is present with probability p_2 independently. Therefore, the probability that v gets attached to G is $1-(1-p_2)^{\binom{n_1}{d-1}}$. In fact, these events occur independently for all $v \in \mathcal{W}$. Consequently,

$$\mathcal{S}_G^{\mathrm{iso}} = \mathrm{Bin}\left(|\mathcal{W}|, 1-(1-p_2)^{\binom{n_1}{d-1}}\right),\ \mu_{\mathrm{iso}} = \mathrm{E}(\mathcal{S}_G^{\mathrm{iso}}) = |\mathcal{W}|(1-(1-p_2)^{\binom{n_1}{d-1}}) = \Omega(|\mathcal{W}|), \qquad (25)$$

where the last equality sign follows from the fact that $p_2 \sim \varepsilon p_1 = \Theta(n^{1-d})$.

Hence, $\mathcal{S}_G = \mathcal{S}_G^{\mathrm{big}} + \mathcal{S}_G^{\mathrm{iso}}$ features a contribution that satisfies a local limit theorem, namely the binomially distributed $\mathcal{S}_G^{\mathrm{iso}}$. Thus, to establish the locality of $\mathcal{S}_G$ (i.e., Lemma 12), we are going to prove that $\mathcal{S}_G$ "inherits" the locality of $\mathcal{S}_G^{\mathrm{iso}}$. To this end, we need to bound $|\mathcal{W}|$, thereby estimating $\mu_{\mathrm{iso}} = \mathrm{E}(\mathcal{S}_G^{\mathrm{iso}})$.

Lemma 15. *We have* $\mathrm{P}\left[|\mathcal{W}| \geq \frac{1}{2}(n-n_1)\exp(-c)\right] \geq 1 - n^{-10}$.

The proof of Lemma 15 is just a standard application of Azuma's inequality, cf. Section 4.6.

Further, let M be the set of all triples (H, F_1, F_2) such that

M1. $\mathrm{P}\left[\mathcal{S}_G = s | H_{2,G} = H,\ \mathcal{F}_1 = F_1,\ \mathcal{F}_2 = F_2\right] \geq n^{-11}$, and
M2. given that $H_{2,G} = H$, $\mathcal{F}_1 = F_1$, and $\mathcal{F}_2 = F_2$, the set $\mathcal{W}$ has size $\geq \frac{1}{2}(n-n_1)\exp(-c) = \Omega(n)$.

Lemma 16. *If* $|s-t| \leq \beta\sqrt{n}$ *for some small enough* $\beta = \beta(\alpha) > 0$, *then* $\mathrm{P}\left[\mathcal{S}_G = t | (H_{2,G}, \mathcal{F}_1, \mathcal{F}_2) \in M\right] \geq (1-\alpha)\mathrm{P}\left[\mathcal{S}_G = s | (H_{2,G}, \mathcal{F}_1, \mathcal{F}_2) \in M\right]$.

Proof. Let $(H, F_1, F_2) \in M$, and let b be the value of $\mathcal{S}_G^{\mathrm{big}}$ given that $H_{2,G} = H$, $\mathcal{F}_1 = F_1$ and $\mathcal{F}_2 = F_2$. Then given that this event occurs, we have $\mathcal{S}_G = s$ iff $\mathcal{S}_G^{\mathrm{iso}} = s - b$. As $(H, F_1, F_2) \in M$, we conclude that

$$\mathrm{P}\left[\mathcal{S}_G = s | H_{2,G} = H,\ \mathcal{F}_1 = F_1,\ \mathcal{F}_2 = F_2\right] = \mathrm{P}\left[\mathrm{Bin}\left(|\mathcal{W}|, 1-(1-p_2)^{\binom{n_1}{d-1}}\right) = s - b\right] \overset{\mathbf{M1}}{\geq} n^{-11}.$$

Therefore, the Chernoff bound (7) implies that $|s - b - \mu_{\mathrm{iso}}| \leq n^{0.6}$. Furthermore, since we assume that $|t-s| \leq \beta n^{1/2}$ for some small $\beta = \beta(\alpha) > 0$ and as $\mu_{\mathrm{iso}} = |\mathcal{W}|(1-(1-p_2)^{\binom{n_1}{d-1}}) \geq \Omega(n)$ due to **M2**, Proposition 3 entails that

$$\mathrm{P}\left[\mathrm{Bin}\left(|\mathcal{W}|, 1-(1-p_2)^{\binom{n_1}{d-1}}\right) = t - b\right] \geq (1-\alpha)\mathrm{P}\left[\mathrm{Bin}\left(|\mathcal{W}|, 1-(1-p_2)^{\binom{n_1}{d-1}}\right) = s - b\right].$$

Thus, the assertion follows from (25). □

Proof of Lemma 12. By Lemmas 15 and 16, we have

$$\begin{aligned}\mathrm{P}\left[\mathcal{S}_G = s\right] &\leq \mathrm{P}\left[\mathcal{S}_G = s | (H_{2,G}, \mathcal{F}_1, \mathcal{F}_2) \notin M\right] \mathrm{P}\left[(H_{2,G}, \mathcal{F}_1, \mathcal{F}_2) \notin M\right] + (1-\alpha)^{-1}\mathrm{P}\left[\mathcal{S}_G = t\right] \\ &\overset{\mathbf{M1,M2}}{\leq} n^{-11} + \mathrm{P}\left[|\mathcal{W}| = o(n)\right] + (1-\alpha)^{-1}\mathrm{P}\left[\mathcal{S}_G = t\right] \leq (1-\alpha)^{-1}\mathrm{P}\left[\mathcal{S}_G = t\right] + n^{-10},\end{aligned}$$

as claimed. □

4.3 Proof of Lemma 9

Let $\mathcal{L}_G$ signify the event that G is the largest component of H_1. Given that $\mathcal{L}_G$ occurs, the edges in $H_3 - G$ do not occur independently anymore. For if $\mathcal{L}_G$ occurs, then $H_1 - G$ does not contain a component on more than $|G|$ vertices. Nonetheless, the following lemma shows that if $E \subset \mathcal{E}(V) \setminus \mathcal{E}(G)$ is a set of edges such that the hypergraph $H(E) = (V, E \cap \mathcal{E}(V \setminus G))$ does not feature a "big" component, then the dependence of the edges is very small. In other words, the probability that the edges E are present in H_3 is very close to the probability that these edges are present in the "artificial" model $H_{3,G}$, in which edges occur independently.

Lemma 17. *For any set $E \subset \mathcal{E}(V) \setminus \mathcal{E}(G)$ such that $\mathcal{N}(H(E)) \le \ln^2 n$ we have*

$$\mathrm{P}\left[E(H_3) \setminus \mathcal{E}(G) = E \,|\, \mathcal{L}_G\right] = (1 + O(n^{-10}))\mathrm{P}\left[E(H_{3,G}) = E\right].$$

Before getting down to the proof of Lemma 17, we first show how it implies Lemma 9. As a first step, we derive that it is actually quite unlikely that either $H_3 - G$ or $H_{3,G} - G$ features a component on $\ge \ln^2 n$ vertices.

Corollary 18. *We have* $\mathrm{P}\left[\mathcal{N}(H_3 - G) > \ln^2 n | \mathcal{L}_G\right], \mathrm{P}\left[\mathcal{N}(H_{3,G} - G) > \ln^2 n\right] = O(n^{-10})$.

Proof. Theorem 4 implies that $\mathrm{P}\left[\mathcal{N}(H_{3,G} - G) > \ln^2 n\right] = O(n^{-10})$, because $H_{3,G}$ simply is a random hypergraph $H_d(n - n_1, p)$, and $\binom{n-n_1}{d-1}p \sim \binom{n-\mu_1}{d-1}p < (d-1)^{-1}$ by (8). Hence, Lemma 17 yields that $\mathrm{P}\left[\mathcal{N}(H_3 - G) \le \ln^2 n | \mathcal{L}_G\right] \ge (1 - O(n^{-10}))\mathrm{P}\left[\mathcal{N}(H_{3,G} - G) \le \ln^2 n\right] \ge 1 - O(n^{-10})$. □

Proof of Lemma 9. Let $\mathcal{A}_s$ denote the set of all subsets $E \subset \mathcal{E}(V) \setminus \mathcal{E}(G)$ such that in the hypergraph (V, E) exactly s vertices in $V \setminus G$ are reachable from G. Moreover, let $\mathcal{B}_s$ signify the set of all $E \in \mathcal{A}_s$ such that $\mathcal{N}(H(E)) \le \ln^2 n$. Then

$$\mathrm{P}\left[\mathcal{S} = s | \mathcal{L}_G\right] = \mathrm{P}\left[E(H_3) \setminus \mathcal{E}(G) \in \mathcal{A}_s | \mathcal{L}_G\right], \text{ and } \mathrm{P}\left[\mathcal{S}_G = s\right] = \mathrm{P}\left[E(H_{3,G}) \in \mathcal{A}_s\right]. \tag{26}$$

Furthermore, by Corollary 18

$$\mathrm{P}\left[E(H_3) \setminus \mathcal{E}(G) \in \mathcal{A}_s \setminus \mathcal{B}_s | \mathcal{L}_G\right] \le \mathrm{P}\left[\mathcal{N}(H_3 - G) > \ln^2 n | \mathcal{L}_G\right] = O(n^{-10}), \tag{27}$$

$$\mathrm{P}\left[E(H_{3,G}) \in \mathcal{A}_s \setminus \mathcal{B}_s\right] \le \mathrm{P}\left[\mathcal{N}(H_{3,G} - G) > \ln^2 n\right] = O(n^{-10}). \tag{28}$$

Combining (26), (27), and (28), we conclude that

$$\begin{aligned} \mathrm{P}\left[\mathcal{S} = s | \mathcal{L}_G\right] &= \mathrm{P}\left[E(H_3) \setminus \mathcal{E}(G) \in \mathcal{B}_s | \mathcal{L}_G\right] + O(n^{-10}) \\ &\overset{\text{Lemma 17}}{=} \mathrm{P}\left[E(H_{3,G}) \in \mathcal{B}_s\right] + O(n^{-10}) = \mathrm{P}\left[\mathcal{S}_G = s\right] + O(n^{-10}), \end{aligned}$$

thereby completing the proof. □

Thus, the remaining task is to prove Lemma 17. To this end, let $\mathcal{H}_1(E)$ denote the event that $\mathcal{E}(V \setminus G) \cap E(H_1) = E$. Moreover, let $\mathcal{H}_2(E)$ signify the event that $\mathcal{E}(V \setminus G) \cap E(H_2) \setminus E(H_1) = E$ (i.e., E is the set of edges added during **R2**). Further, let $\mathcal{H}_3(E)$ be the event that $\mathcal{E}(G, V \setminus G) \cap E(H_3) = E$ (i.e., E consists of all edges added by **R3**). In addition, define events $\mathcal{H}_{1,G}(E)$, $\mathcal{H}_{2,G}(E)$, $\mathcal{H}_{3,G}(E)$ analogously, with H_1, H_2, H_3 replaced by $H_{1,G}$, $H_{2,G}$, $H_{3,G}$. Finally, let $\mathcal{C}_G$ denote the event that G is a component of H_1. In order to prove Lemma 17, we establish the following.

Lemma 19. *Let $E_1 \subset \mathcal{E}(V \setminus G)$, $E_2 \subset \mathcal{E}(V \setminus G) \setminus E_1$, and $E_3 \subset \mathcal{E}(G, V \setminus G)$. Moreover, suppose that $\mathcal{N}(H(E_1)) \le \ln^2 n$. Then* $\mathrm{P}\left[\bigwedge_{i=1}^3 \mathcal{H}_i(E_i) | \mathcal{L}_G\right] = (1 + O(n^{-10}))\mathrm{P}\left[\bigwedge_{i=1}^3 \mathcal{H}_{i,G}(E_i)\right]$.

Proof. Clearly,

$$\mathrm{P}\left[\bigwedge_{i=1}^3 \mathcal{H}_i(E_i) | \mathcal{L}_G\right] = \frac{\mathrm{P}\left[\mathcal{H}_2(E_2) \wedge \mathcal{H}_3(E_3) | \mathcal{L}_G \wedge \mathcal{H}_1(E_1)\right] \mathrm{P}\left[\mathcal{H}_1(E_1) \wedge \mathcal{L}_G\right]}{\mathrm{P}\left[\mathcal{L}_G\right]}. \tag{29}$$

Furthermore, since **R2** and **R3** add edges independently of the 1st round with probability p_2, and because the same happens during **R2'** and **R3'**, we have

$$\mathrm{P}\left[\mathcal{H}_2(E_2) \wedge \mathcal{H}_3(E_3) | \mathcal{L}_G \wedge \mathcal{H}_1(E_1)\right] = \mathrm{P}\left[\mathcal{H}_{2,G}(E_2) \wedge \mathcal{H}_{3,G}(E_3) | \mathcal{H}_{1,G}(E_1)\right]. \tag{30}$$

Moreover, given that $\mathcal{H}_1(E_1)$ occurs, $H_1 - G$ has no component on more than $\ln^2 n$ vertices. Hence, G is the largest component of H_1 iff G is a component; that is, given that $\mathcal{H}_1(E_1)$ occurs, the events $\mathcal{L}_G$ and $\mathcal{C}_G$ are equivalent. Therefore, $\mathrm{P}\left[\mathcal{L}_G \wedge \mathcal{H}_1(E_1)\right] = \mathrm{P}\left[\mathcal{C}_G \wedge \mathcal{H}_1(E_1)\right]$. Further, whether or not G is a component of H_1 is independent of the edges contained in $V \setminus G$, and thus $\mathrm{P}\left[\mathcal{C}_G \wedge \mathcal{H}_1(E_1)\right] =$

$\mathrm{P}\left[\mathcal{C}_G\right]\mathrm{P}\left[\mathcal{H}_1(E_1)\right]$. Hence, as each edge in E_1 is present in H_1 as well as in $H_{1,G}$ with probability p_1 independently, we obtain

$$\mathrm{P}\left[\mathcal{L}_G \wedge \mathcal{H}_1(E_1)\right] = \mathrm{P}\left[\mathcal{C}_G\right] p_1^{|E_1|}(1-p_1)^{\mathcal{E}(V\setminus G)-|E_1|} = \mathrm{P}\left[\mathcal{C}_G\right]\mathrm{P}\left[\mathcal{H}_{1,G}(E_1)\right]. \tag{31}$$

Combining (29), (30), and (31), we obtain

$$\mathrm{P}\left[\bigwedge_{i=1}^{3}\mathcal{H}_i(E_i)|\mathcal{L}_G\right] = \frac{\mathrm{P}\left[\mathcal{C}_G\right]}{\mathrm{P}\left[\mathcal{L}_G\right]}\cdot\mathrm{P}\left[\bigwedge_{i=1}^{3}\mathcal{H}_{i,G}(E_i)\right]. \tag{32}$$

Since by Theorem 4 with probability $\geq 1-n^{-10}$ the random hypergraph $H_1 = H_d(n,p_1)$ has precisely one component of order $\Omega(n)$, we get $\frac{\mathrm{P}[\mathcal{C}_G]}{\mathrm{P}[\mathcal{L}_G]} = 1 + O(n^{-10})$. Hence, (32) implies the assertion. □

Proof of Lemma 17. For any set $E \subset \mathcal{E}(V)\setminus\mathcal{E}(G)$ let $\mathcal{F}(E)$ denote the set of all decompositions (E_1,E_2,E_3) of E into three disjoint sets such that $E_1,E_2 \subset \mathcal{E}(V\setminus G)$ and $E_3 \subset \mathcal{E}(G,V\setminus G)$. If $\mathcal{N}(H(e)) \leq \ln^2 n$, then Lemma 19 implies that

$$\begin{aligned}\mathrm{P}\left[E(H_3)\setminus\mathcal{E}(G) = E|\mathcal{L}_G\right] &= \sum_{(E_1,E_2,E_3)\in\mathcal{F}(E)}\mathrm{P}\left[\bigwedge_{i=1}^{3}\mathcal{H}_i(E_i)|\mathcal{L}_G\right]\\ &= (1+O(n^{-10}))\sum_{(E_1,E_2,E_3)\in\mathcal{F}(E)}\mathrm{P}\left[\bigwedge_{i=1}^{3}\mathcal{H}_{i,G}(E_i)\right] = (1+O(n^{-10}))\mathrm{P}\left[E(H_{3,G}) = E\right],\end{aligned}$$

as claimed. □

4.4 Proof of Lemma 10

Recall that S_G signifies the set of all vertices $v \in V\setminus G$ that are reachable from G in $H_{3,G}$, so that $\mathcal{S}_G = |S_G|$. Letting $\mathcal{C}_v$ denote the component of $H_{2,G}$ that contains $v \in V$, we have

$$\mathrm{E}(\mathcal{S}_G) = \sum_{v\in V\setminus G}\mathrm{P}\left[v\in S_G\right] = \sum_{v\in V\setminus G}\sum_{k=1}^{n-n_1}\mathrm{P}\left[v\in S_G||\mathcal{C}_v| = k\right]\mathrm{P}\left[|\mathcal{C}_v| = k\right] \tag{33}$$

Since $H_{2,G}$ is just a random hypergraph $H_d(n-n_1,p)$, and because $\binom{n-n_1}{d-1}p \sim \binom{n-\mu_1}{d-1}p < (d-1)^{-1}$ by (8), Theorem 4 entails that $\mathcal{N}(H_{2,G}) \leq \ln^2 n$ with probability $\geq 1-n^{-10}$. Therefore, (33) yields

$$\mathrm{E}(\mathcal{S}_G) = o(1) + \sum_{v\in V\setminus G}\sum_{1\leq k\leq \ln^2 n}\mathrm{P}\left[v\in S_G||\mathcal{C}_v| = k\right]\mathrm{P}\left[|\mathcal{C}_v| = k\right]. \tag{34}$$

To estimate $\mathrm{P}\left[v\in S_G||\mathcal{C}_v| = k\right]$, let $z = z(n_1) = (n_1-\mu_1)/\sigma_1$, $\xi_0 = \exp\left[-p_2\left[\binom{n-1}{d-1} - \binom{n-\mu_1}{d-1}\right]\right]$, and $\xi(z) = \xi_0\left[1 + z\sigma_1 p_2\binom{n-\mu_1}{d-2}\right]$. Additionally, let $\zeta(z) = \binom{n-n_1}{d-1}p \sim \binom{n-\mu_1}{d-1}p - z\sigma_1\binom{n-\mu_1}{d-2}p$.

Lemma 20. *For all* $1 \leq k \leq \ln^2 n$ *we have* $\mathrm{P}\left[v\in S_G \mid |\mathcal{C}_v| = k\right] = 1 - \xi(z)^k + O(n^{-1}\cdot\mathrm{polylog}\, n)$.

Proof. Suppose that $|\mathcal{C}_v| = k$ but $v \notin S_G$. This is the case iff in $H_{3,G}$ there occurs no edge that is incident to both G and $\mathcal{C}_v$. Letting $\mathcal{E}(G,\mathcal{C}(v))$ denote the set of all possible edges connecting G and $\mathcal{C}_v$, we shall prove below that

$$\begin{aligned}|\mathcal{E}(G,\mathcal{C}_v)| &= k\left[\binom{n}{d-1} - \binom{n-\mu_1}{d-1} + \frac{z\sigma_1}{d-1}\binom{n-\mu_1}{d-2}\right] + O(n^{d-2}\cdot\mathrm{polylog}\, n)\\ &= O(n^{d-1}\cdot\mathrm{polylog}\, n)\end{aligned} \tag{35}$$

By construction every edge in $\mathcal{E}(G,\mathcal{C}_v)$ occurs in $H_{3,G}$ with probability p_2 independently. Therefore,

$$\begin{aligned}\mathrm{P}\left[v \notin S_G \| \mathcal{C}_v| = k\right] \;&=\; (1-p_2)^{|\mathcal{E}(G,\mathcal{C}_v)|} = (1+O(n^{-1}\cdot \operatorname{polylog} n))\exp\left[-p_2|\mathcal{E}(G,\mathcal{C}_v)|\right] \\ &\overset{(35)}{=} (1+O(n^{-1}\cdot \operatorname{polylog} n))\xi(z)^k,\end{aligned}$$

hence the assertion follows.

Thus, the remaining task is to prove (35). As a first step, we show that

$$|\mathcal{E}(G,\mathcal{C}_v)| = \binom{n}{d} - \binom{n-k}{d} - \binom{n-n_1}{d} + \binom{n-n_1-k}{d}. \tag{36}$$

For there are $\binom{n}{d}$ possible edges in total, among which $\binom{n-k}{d}$ contain no vertex of $\mathcal{C}_v$, $\binom{n-n_1}{d}$ contain no vertex of G, and $\binom{n-n_1-k}{d}$ contain neither a vertex of $\mathcal{C}_v$ nor of G; thus, (36) follows from the inclusion/exclusion formula. Furthermore, as $k = O(\operatorname{polylog} n)$, we have $\binom{n}{d} - \binom{n-k}{d} = (1+O(n^{-1}\cdot \operatorname{polylog} n))k\binom{n}{d-1}$ and $\binom{n-n_1}{d} - \binom{n-n_1-k}{d} = (1+O(n^{-1}\cdot \operatorname{polylog} n))k\binom{n-n_1}{d-1}$. Thus (36) yields

$$|\mathcal{E}(G,\mathcal{C}(v))| = (1+O(n^{-1}\cdot \operatorname{polylog} n))k\left[\binom{n}{d-1} - \binom{n-n_1}{d-1}\right]. \tag{37}$$

As $n_1 = \mu_1 + z\sigma_1$, we have $\binom{n-n_1}{d-1} = \binom{n-\mu_1}{d-1} - z\sigma_1\binom{n-n_1}{d-2} + O(n^{d-2}\cdot \operatorname{polylog} n)$, so that (35) follows from (37). $\square$

Let $q(\zeta,\xi) = \sum_{k=1}^{\infty} q_k(\zeta)\xi^k$ be the function from Proposition 5. Combining (34) with Proposition 5 and Lemma 20, we conclude that

$$\mathrm{E}(\mathcal{S}_G) = o(n^{1/2}) + q((n-n_1)p,\xi(z))(n-n_1) = o(n^{1/2}) + q(\zeta(z),\xi(z))(n-n_1). \tag{38}$$

Since q is differentiable (cf. Proposition 5), we let $\Delta_\zeta = \frac{\partial q}{\partial \zeta}(\zeta(0),\xi(0))$ and $\Delta_\xi = \frac{\partial q}{\partial \xi}(\zeta(0),\xi(0))$. As $\zeta(z)-\zeta(0), \xi(z)-\xi(0) = O(n^{-1/2})$, we get

$$\begin{aligned}q(\zeta(z),\xi(z)) - q(\zeta(0),\xi(0)) &= (\zeta(z)-\zeta(0))\Delta_\zeta + (\xi(z)-\xi(0))\Delta_\xi + o(n^{-1/2}) \\ &= z\sigma_1\binom{n-\mu_1}{d-2}\left[\xi_0\Delta_\xi p_2 - \Delta_\zeta p\right] + o(n^{-1/2}).\end{aligned} \tag{39}$$

Finally, let $\mu_{\mathcal{S}} = (n-\mu_1)q(\zeta(0),\xi(0))$ and $\lambda_{\mathcal{S}} = q(\zeta(0),\xi(0)) - (d-1)\left[\varepsilon\xi_0\Delta_\xi - \Delta_\zeta\right]\binom{n-\mu_1}{d-1}p$. Then combining (38) and (39), we see that $\mathrm{E}(\mathcal{S}_G) = \mu_{\mathcal{S}} + z\sigma_1\lambda_{\mathcal{S}} + o(\sqrt{n})$, as desired.

4.5 Proof of Lemma 11

Remember that S_G denotes the set of all "attached" vertices, and $N_{v,G}$ the order of the component of $v \in V \setminus G$ in the graph $H_{2,G}$.

The following lemma provides an asymptotic formula for $\mathrm{Var}(\mathcal{S}_G)$.

Lemma 21. *Let $r_{G,i} = \mathrm{P}\left[N_{v,G} = i \wedge v \in S_G\right]$ and $\bar{r}_{G,i} = \mathrm{P}\left[N_{v,G} = i \wedge v \notin S_G\right]$ for any vertex $v \in V \setminus G$. Moreover, set $r_G = \sum_{i=1}^{L} r_{G,i}$, $R_G = \sum_{i=1}^{L} i r_{G,i}$, $\bar{R}_G = \sum_{i=1}^{L} i\bar{r}_{G,i}$ for $L = \lceil \ln^2 n \rceil$. In addition, let $\alpha_G = 1 - |G|/n$ and*

$$\Gamma_G = (1-R_G)(R_G - r_G) + ((d-1)c-1)\frac{R_G^2}{r_G} + R_G + (d-1)(1-\alpha_G^{d-2})\varepsilon c\bar{R}_G^2 + \frac{1-\alpha_G^{d-2}}{1-\alpha_G^{d-1}}\bar{R}_G. \tag{40}$$

Then $\mathrm{Var}(\mathcal{S}_G) \sim \alpha_G^2\Gamma_G n + \alpha_G r_G(1-r_G)n$.

Before we get down to the proof of Lemma 21, we observe that it implies Lemma 11.

Proof of Lemma 11. By Theorem 4 part 2 together with Lemma 9 we know that with probability at least $1 - n^{-8}$ there are no components of order $> \ln^2 n$ inside of $V \setminus G$. Let $q(\zeta, \xi) = \sum_{k=1}^{\infty} q_k(\zeta)\xi^k$ be the function from Proposition 5, and let $\xi(z)$ be as in Lemma 20. Then Proposition 5 and Lemma 20 entail that for all $v \in V \setminus G$

$$r_{G,i} = q_i\left(\binom{n-|G|}{d-1}p\right)\xi((|G|-\mu_1)/\sigma_1), \;\; \bar{r}_{G,i} \sim q_i\left(\binom{n-|G|}{d-1}p\right)(1-\xi((|G|-\mu_1)/\sigma_1)).$$

By (9) there exists a number $0 < \gamma < 1$ such that $q_i\left(\binom{n-|G|}{d-1}p\right) \leq \gamma^i$. Since $0 \leq \xi((|G|-\mu_1)/\sigma_1) \leq 1$, this yields $r_{G,i}, \bar{r}_{G,i} \leq \gamma^i$. Hence, $R_G, \bar{R}_G = O(1)$, so that Lemma 21 implies $\mathrm{Var}(\mathcal{S}_G) = O(n)$.

Finally, if $G' \subset V$ satisfies $||G'|-|G|| \leq n^{0.9}$, then $|\binom{n-|G|}{d-1}p - \binom{n-|G'|}{d-1}p| = O(|G|-|G'|)/n$, because $p = O(n^{1-d})$. Therefore, $|q_i\left(\binom{n-|G|}{d-1}p\right) - q_i\left(\binom{n-|G'|}{d-1}p\right)| = O(|G|-|G'|)/n$, because the function $\zeta \mapsto q_i(\zeta)$ is differentiable. Similarly, as $\xi(z) = \xi_0(1 + z\sigma_1 p_2\binom{n-\mu_1}{d-2})$ for some fixed $\xi_0 = \Theta(1)$, we have $|\xi((|G|-\mu_1)/\sigma_1) - \xi((|G'|-\mu_1)/\sigma_1)| = O(|G|-|G'|)/n$. Consequently, $|r_{G,i} - r_{G',i}| = O(|G|-|G'|)/n$ and $|\bar{r}_{G,i} - \bar{r}_{G',i}| = O(|G|-|G'|)/n$, and thus

$$|r_G - r_{G'}|, |R_G - R_{G'}|, |\bar{R}_G - \bar{R}_{G'}| = O(|G|-|G'|)/n = O(n^{-0.1}).$$

Hence, Lemma 21 implies that $|\mathrm{Var}(\mathcal{S}_G) - \mathrm{Var}(\mathcal{S}_{G'})| = o(n)$. □

The remaining task is to establish Lemma 21. Keeping G fixed, in the sequel we constantly omit the subscript G in order to ease up the notation; thus, we write α instead of α_G etc. As a first step, we compute $\mathrm{P}(v, w \in S) - r^2$. Setting

$$\begin{aligned}
S_1 &= \sum_{i,j=1}^{L} \left[\mathrm{P}\left[N_w = j \wedge w \in S | w \notin C_v, N_v = i, v \in S\right] - \mathrm{P}\left[N_w = j \wedge w \in S\right]\right] \\
&\qquad\qquad \times \mathrm{P}\left[w \notin C_v | N_v = i, v \in S\right] \mathrm{P}\left[N_v = i \wedge v \in S\right], \\
S_2 &= (1-r)\sum_{i=1}^{L} \mathrm{P}\left[w \in C_v | N_v = i, v \in S\right] \mathrm{P}\left[N_v = i \wedge v \in S\right],
\end{aligned}$$

we have $\mathrm{P}(v, w \in S) - r^2 = S_1 + S_2$.

To compute S_2, observe that whether $w \in C_v$ depends only on N_v, but not on the event $v \in S$. Therefore, $\mathrm{P}\left[w \in C_v | N_v = i, v \in S\right] = \mathrm{P}\left[w \in C_v | N_v = i\right] = \binom{n-2}{i-2}\binom{n-1}{i-1}^{-1} = \frac{i-1}{n-1}$, because given that $N_v = i$, there are $\binom{n_0-1}{i-1}$ ways to choose the set $C_v \subset V \setminus G$, while there are $\binom{n_0-2}{i-2}$ ways to choose C_v in such a way that $w \in C_v$. As a consequence,

$$S_2 \sim \frac{1-r}{n-1}\sum_{i=1}^{L}(i-1)\mathrm{P}\left[N_v = i \wedge v \in S\right] = \frac{1-r}{n-1}(R-r).$$

With respect to S_1, we let

$$\begin{aligned}
P_1(i,j) &= \mathrm{P}\left[N_w = j | w \notin C_v, N_v = i\right], \\
P_2(i,j) &= \mathrm{P}\left[w \in S | N_w = j, w \notin C_v, N_v = i, v \in S\right],
\end{aligned}$$

so that

$$\begin{aligned}
S_1 &= \sum_{i,j}\left[P_1(i,j)P_2(i,j) - \mathrm{P}\left[N_w = j \wedge w \in S\right]\right]\mathrm{P}\left[w \notin C_v | N_v = i, v \in S\right]\mathrm{P}\left[N_v = i \wedge v \in S\right] \\
&\sim \sum_{i,j}\left[P_1(i,j)P_2(i,j) - \mathrm{P}\left[N_w = j\right]\mathrm{P}\left[w \in S | N_w = j\right]\right]\mathrm{P}\left[N_v = i \wedge v \in S\right].
\end{aligned}$$

Lemma 22. *We have* $P_1(i,j)\mathrm{P}\left[N_w=j\right]^{-1} = 1+\frac{((d-1)c-1)ij+i}{n-n_1}+O(n^{-2}\cdot\mathrm{polylog}\,n)$.

Proof. This argument is similar to the one used in the proof of Lemma 41 in [7]. Remember that if we restrict our view on $H_{3,G}$ to the set $V\setminus G$ the hypergraph is similar to a $H_d(n-n_1,p)$. In order to estimate S_1, we observe that

$$\mathrm{P}\left[N_w=j \text{ in } H_d(n-n_1,p)\,|\,N_v=i,\, w\notin C_v\right]=\mathrm{P}\left[N_w=j \text{ in } H_d(n-n_1,p)\setminus C_v\right]. \tag{41}$$

Given that $N_v=i$, $H_d(n,p)\setminus C_v$ is distributed as a random hypergraph $H_d(n-n_1-i,p)$. Hence, the probability that $N_w=j$ in $H_d(n,p)\setminus C_v$ equals the probability that a given vertex of $H_d(n-n_1-i,p)$ belongs to a component of order j. Therefore, we can compare $\mathrm{P}\left[N_w=j \text{ in } H_d(n-n_1,p)\setminus C_v\right]$ and $\mathrm{P}\left[N_w=j \text{ in } H_d(n-n_1,p)\right]$ as follows: in $H_d(n-n_1-i,p)$ there are $\binom{n-n_1-i-1}{j-1}$ ways to choose the set $C_w\setminus\{j\}$. Moreover, there are $\binom{n-n_1-i}{d}-\binom{n-n_1-i-j}{d}-\binom{j}{d}$ possible edges connecting the chosen set C_w with $V\setminus C_w$, and as C_w is a component, none of these edges is present. Since each such edge is present with probability p independently, the probability that there is no C_w-$V\setminus C_w$ edge equals

$$(1-p)^{\binom{n-n_1-i}{d}-\binom{n-n_1-i-j}{d}-\binom{j}{d}}.$$

By comparison, in $H_d(n-n_1,p)$ there are $\binom{n-n_1-1}{j-1}$ ways to choose the vertex set of C_w. Further, there are $\binom{n-n_1}{d}-\binom{n-n_1-j}{d}-\binom{j}{d}$ possible edges connecting C_w and $V\setminus C_w$, each of which is present with probability p independently. Thus, letting $\gamma=\binom{n-n_1-i}{d}-\binom{n-n_1-i-j}{d}-\left[\binom{n-n_1}{d}-\binom{n-n_1-j}{d}\right]$, we obtain

$$\frac{\mathrm{P}\left[N_w=j \text{ in } H_d(n-n_1,p)\setminus C_v\right]}{\mathrm{P}\left[N_w=j \text{ in } H_d(n-n_1,p)\right]}=\binom{n-n_1-i-1}{j-1}\binom{n-n_1-1}{j-1}^{-1}(1-p)^{\gamma}. \tag{42}$$

Concerning the quotient of the binomial coefficients, we have

$$\binom{n-n_1-i-1}{j-1}\binom{n-n_1-1}{j-1}^{-1}=\exp\left[-\frac{i(j-1)}{n-n_1}+O(n^{-2}\cdot\mathrm{polylog}\,n)\right]. \tag{43}$$

Moreover, $\gamma=\binom{n-n_1}{d}\left[\frac{(n-n_1-i)_d+(n-n_1-j)_d-(n-n_1-i-j)_d}{(n-n_1)_d}-1\right]$. Expanding the falling factorials, we get

$$\gamma=\binom{n-n_1}{d}\left[\frac{\binom{d}{2}(i^2+j^2-(i+j)^2)}{(n-n_1)^2}+O(n^{-3}\cdot\mathrm{polylog}\,n)\right]=-\binom{n-n_1}{d-2}ij+O(n^{d-3}\cdot\mathrm{polylog}\,n). \tag{44}$$

Plugging (43) and (44) into (42), we obtain

$$\begin{aligned}\frac{\mathrm{P}\left[N_w=j \text{ in } H_d(n-n_1,p)\setminus C_v\right]}{\mathrm{P}\left[N_w=j \text{ in } H_d(n-n_1,p)\right]}&=\exp\left[-\frac{i(j-1)}{n-n_1}+O(n^{-2}\cdot\mathrm{polylog}\,n)\right](1-p)^{-\binom{n-n_1}{d-2}ij+O(n^{d-3}\cdot\mathrm{polylog}\,n)}\\&=\exp\left[-\frac{i(j-1)}{n-n_1}+\binom{n-n_1}{d-2}ijp+O(n^{-2}\cdot\mathrm{polylog}\,n)\right]\\&=1+(n-n_1)^{-1}\left[((d-1)c-1)ij+i\right]+O(n^{-2}\cdot\mathrm{polylog}\,n).\end{aligned}$$

Therefore, by (41)

$$\begin{aligned}&\mathrm{P}\left[N_w=j|N_v=i,\,w\notin C_v\right]-\mathrm{P}\left[N_w=j \text{ in } H_d(n-n_1,p)\right]\\&\quad=\mathrm{P}\left[N_w=j \text{ in } H_d(n-n_1,p)\right]\left[n^{-1}\left[((d-1)c-1)ij+i\right]+O(n^{-2}\cdot\mathrm{polylog}\,n)\right].\end{aligned} \tag{45}$$

□

Lemma 23. *Setting* $\gamma_1=\frac{1-\alpha^{d-2}}{\mathrm{P}[v\in S|N_v=i](1-\alpha^{d-1})}$ *and* $\gamma_2=(d-1)(1-\alpha^{d-2})\varepsilon c$*, we have* $P_2(i,j)-\mathrm{P}\left[w\in S|N_w=j\right]=n^{-1}\mathrm{P}\left[w\notin S|N_w=j\right](j\gamma_1-ij\gamma_2)+O(n^{-2}\cdot\mathrm{polylog}\,n)$.

Proof. Let $\mathcal{F}$ be the event that $N_w = j$, $w \notin C_v$, $N_v = i$, and $v \in S$. Moreover, let $\mathcal{Q}$ be the event that in H_3 there exists an edge incident to the three sets C_v, C_w, and G simultaneously, so that $P_2(i,j) = \mathrm{P}[\mathcal{Q}|\mathcal{F}] + \mathrm{P}[w \in S|\neg\mathcal{Q},\mathcal{F}]\,\mathrm{P}[\neg\mathcal{Q}|\mathcal{F}]$.

To bound $\mathrm{P}[w \in S|\neg\mathcal{Q},\mathcal{F}] - \mathrm{P}[w \in S|N_w = j]$, we condition on the event that C_v and C_w are fixed disjoint sets of sizes i and j. Let Q' signify the probability that C_w is reachable from G in $H_{3,G}$, and let Q denote the probability that C_w is reachable from G in $H_{3,G}$, and that the event $\neg\mathcal{Q}$ occurs. Then Q' corresponds to $\mathrm{P}[w \in S|N_w = j]$ and Q to $\mathrm{P}[w \in S|\neg\mathcal{Q},\mathcal{F}]$, so that our aim is to estimate $Q - Q'$. As there are $|\mathcal{E}(G,C_v)| - |\mathcal{E}(G,C_v,C_w)|$ possible edges that join C_v and G but avoid C_w, each of which is present in $H_{3,G}$ with probability p_2 independently, we have

$$Q = 1 - (1-p_2)^{|\mathcal{E}(G,C_v)| - |\mathcal{E}(G,C_v,C_w)|}, \text{ while } Q' = 1 - (1-p_2)^{|\mathcal{E}(G,C_w)|}.$$

Therefore,

$$\begin{aligned} Q - Q' &= (1-p_2)^{|\mathcal{E}(G,C_w)|}\left[1 - (1-p_2)^{-|\mathcal{E}(G,C_v,C_w)|}\right] \\ &\sim (1-Q')\left(1 - \exp\left[p_2|\mathcal{E}(G,C_v,C_w)|\right]\right) \sim ij(Q'-1)\left[\binom{n}{d-2} - \binom{n_0}{d-2}\right]p_2. \end{aligned}$$

As $\binom{n-1}{d-1}p_2 \sim \varepsilon c$, we thus get

$$\mathrm{P}[w \in \mathcal{S}|\neg\mathcal{Q},\mathcal{F}] - \mathrm{P}[w \in S|N_w = j] \sim ij(\mathrm{P}[w \in S|N_w = j] - 1)(d-1)(1-\alpha^{d-2})\varepsilon c n^{-1}. \quad (46)$$

With respect to $\mathrm{P}[\mathcal{Q}|\mathcal{F}]$, we let $\mathcal{K}$ signify the number of edges joining C_v and G. Given that $\mathcal{F}$ occurs, $\mathcal{K}$ is asymptotically Poisson with mean $\lambda_i = i\left[\binom{n}{d-1} - \binom{n_0}{d-1}\right]p_2 \sim i(1-\alpha^{d-1})\varepsilon c$. Moreover, given that $\mathcal{K} = k$, the probability that one of these k edges hits C_w is $\mathcal{P}(k) \sim \frac{k\mathcal{E}(G,C_v,C_w)}{\mathcal{E}(C_v,G)}$, and thus

$$\mathcal{P}(k) \sim jk\left[\binom{n}{d-2} - \binom{n_0}{d-2}\right]\left[\binom{n}{d-1} - \binom{n_0}{d-1}\right]^{-1} \sim jk(d-1)\frac{1-\alpha^{d-2}}{1-\alpha^{d-1}}.$$

Consequently,

$$\mathrm{P}[\mathcal{Q}|\mathcal{F}] \sim \frac{\exp(-\lambda_i)}{1-\exp(-\lambda_i)}\sum_{k\geq 1}\frac{jk\lambda_i^k}{k!}\mathcal{P}(k) \sim \frac{j(1-\alpha^{d-2})}{n(1-\exp(-\lambda_i))(1-\alpha^{d-1})}. \quad (47)$$

Combining (46) and (47), we obtain the assertion. $\square$

Thus,

$$\begin{aligned} nS_1 &\sim \sum_{i=1}^{L}\mathrm{P}[v \in S \wedge N_v = i] \\ &\qquad \times \sum_{j=1}^{L}\left[((d-1)c-1)ij + i\right]\mathrm{P}[w \in S \wedge N_w = j] + \mathrm{P}[w \notin S \wedge N_w = j]\left[\gamma_1 j + \gamma_2 ij\right] \\ &= ((d-1)c-1)\frac{R^2}{r} + R + \gamma_2\bar{R}^2 + \sum_{i=1}^{N}\frac{1-\alpha^{d-2}}{1-\alpha^{d-1}}\mathrm{P}[N_v = i]\,\bar{R} \\ &= ((d-1)c-1)\frac{R^2}{r} + R + (d-1)(1-\alpha^{d-2})\varepsilon c\bar{R}^2 + \frac{1-\alpha^{d-2}}{1-\alpha^{d-1}}\bar{R}. \end{aligned}$$

Hence, letting $\mathit{\Gamma}$ be as defined by (40) we have $\mathrm{P}[v,w \in S] - \mathrm{P}[v \in S]\,\mathrm{P}[w \in S]) \sim \mathit{\Gamma}/n$. Consequently, $\mathrm{Var}(S) \sim \alpha\mathit{\Gamma} n + \alpha^2 r(1-r)n$.

4.6 Proof of Lemma 15

The probability that a vertex $v \in V \setminus G$ is isolated in $H_{3,G}$ is at least $(1-p)^{\binom{n_1-1}{d-1}}(1-p_2)^{\binom{n}{d-1}} \sim \exp(-p\binom{n_1-1}{d-1} - \varepsilon p\binom{n}{d-1}) \geq \exp(-c)$. Therefore,

$$\mathrm{E}(|\mathcal{W}|) \geq (1-o(1))\exp(-c)(n-n_1). \tag{48}$$

To show that $|\mathcal{W}|$ is concentrated about its mean, we employ the following version of Azuma's inequality (cf. [11, p. 38]).

Lemma 24. *Let $\Omega = \prod_{i=1}^{K} \Omega_i$ be a product of probability spaces. Moreover, let $X : \Omega \to \mathbf{R}$ be a random variable that satisfies the following Lipschitz condition.*

If two tuples $\omega = (\omega_i)_{1 \leq i \leq K}, \omega' = (\omega'_i)_{1 \leq i \leq K} \in \Omega$ differ only in their j'th components for some $1 \leq j \leq K$, then $|X(\omega) - X(\omega')| \leq 1$. (49)

Then $\mathrm{P}\left[|X - \mathrm{E}(X)| \geq t\right] \leq 2\exp(-\frac{t^2}{2K})$, *provided that* $\mathrm{E}(X)$ *exists.*

Using Lemma 24, we shall establish the following.

Corollary 25. *Let Y be a random variable that maps the set of all d-uniform hypergraphs with vertex set V to $[0,n]$. Assume that Y satisfies the following condition.*

Let H be a hypergraph, and let $e \in \mathcal{E}(V)$. Then $|Y(H) - Y(H+e)|, |Y(H) - Y(H-e)| \leq 1$. (50)

Then $\mathrm{P}\left[|Y(H_{3,G}) - \mathrm{E}(Y(H_{3,G}))| \geq n^{0.66}\right] \leq \exp(-n^{0.01})$.

Proof. In order to apply Lemma 24, we need to decompose $H_{3,G}$ into a product $\prod_{i=1}^{K} \Omega_i$ of probability spaces. To this end, consider an arbitrary decomposition of the set $\mathcal{E}(V)$ of all possible edges into sets $\mathcal{E}_1 \cup \cdots \cup \mathcal{E}_K$ so that $K \leq n$ and $\mathrm{E}(E(H_{3,G}) \cap \mathcal{E}_j) \leq n^{0.1}$ for all $1 \leq j \leq K$; such a decomposition exists, because the expected number of edges of $H_{3,G}$ is $\leq \binom{n}{d}p = O(n)$. Now, let Ω_e be a Bernoulli experiment with success probability p for each $e \in \mathcal{E}(V \setminus G)$, resp. with success probability p_2 for $e \in \mathcal{E}(G, V \setminus G)$. Then setting $\Omega_i = \prod_{e \in \mathcal{E}_i} \Omega_e$, we obtain a product decomposition $H_{3,G} = \prod_{i=1}^{K} \Omega_i$.

In addition, construct for each hypergraph H with vertex set V another hypergraph H^* by removing from H all edges $e \in \mathcal{E}_i$ such that $|E(H) \cap \mathcal{E}_i| \geq 4n^{0.1}$ $(1 \leq i \leq K)$. Since $|E(H_{3,G}) \cap \mathcal{E}_i|$ is the sum of two binomially distributed variables, the Chernoff bound (7) implies that $\mathrm{P}\left[|E(H_{3,G}) \cap \mathcal{E}_i|\right] \geq 4n^{0.1}) \leq \exp(-n^{0.05})$. As $K \leq n$, this entails

$$\mathrm{P}\left[H_{3,G} \neq H^*_{3,G}\right] \leq K\exp(-n^{0.05}) \leq \exp(-n^{0.04}), \text{ so that} \tag{51}$$

$$|\mathrm{E}(Y(H_{3,G})) - \mathrm{E}(Y(H^*_{3,G}))| \leq 1 \qquad \text{[because } 0 \leq Y \leq n\text{]}. \tag{52}$$

As a next step, we claim that $Y^*(H) = \frac{1}{4}n^{-0.1}Y(H^*)$ satisfies the Lipschitz condition (49). For by construction modifying (i.e., adding or removing) an arbitrary number of edges belonging to a single factor $\mathcal{E}_i$ can affect at most $4n^{0.1}$ edges of H^*. Hence, (50) implies that $Y^*(H)$ satisfies (49). Therefore, Lemma 24 entails that

$$\mathrm{P}\left[|Y(H^*_{3,G}) - \mathrm{E}(Y(H^*_{3,G}))| \geq n^{0.63}\right] \leq \mathrm{P}\left[|Y^*(H_{3,G}) - \mathrm{E}(Y^*(H_{3,G}))| \geq n^{0.52}\right] \leq \exp(-n^{0.02}). \tag{53}$$

Finally, combining (51), (52), and (53), we conclude that

$$\begin{aligned}\mathrm{P}\left[|Y(H_{3,G}) - \mathrm{E}(Y(H_{3,G}))| \geq n^{0.64}\right] &\leq \mathrm{P}\left[|Y^*(H) - \mathrm{E}(Y^*(H))| \geq n^{0.63}\right] + \mathrm{P}\left[H_{3,G} \neq H^*_{3,G}\right] \\ &\leq \exp(-n^{0.01}),\end{aligned}$$

thereby completing the proof. □

Finally, since $|\mathcal{W}|/d$ satisfies (50), Lemma 15 follows from Corollary 25 and (48).

5 Normality via Stein's Method

In this section we will use Stein's Method to prove that $\mathcal{N}(H_d(n,p))$ as well as $\mathcal{S}_G$ tend (after suitable normalization) in distribution to the normal distribution. This proofs Proposition 6 as well as Theorem 1 and Lemma 14. First we will define a general setting for using Stein's Method with random hypergraphs which defines some conditions the random variables have to fulfill. Then we show in two lemmas (Lemma 28 and Lemma 29) that the random variables corresponding to $\mathcal{N}(H_d(n,p))$ and $\mathcal{S}_G$ do indeed comply to the conditions and last but not least a quite technical part will show how to derive the limiting distribution from the conditions.

5.1 Stein's method for random hypergraphs

Let $\mathcal{E}$ be the set of all subsets of size d of $V=\{1,\dots,n\}$, and let $\mathcal{H}$ be the power set of $\mathcal{E}$. Moreover, let $0\le p_e\le 1$ for each $e\in\mathcal{E}$, and define a probability distribution on $\mathcal{H}$ by letting $\mathrm{P}\left[H\right]=\prod_{e\in H}p_e\cdot\prod_{e\in\mathcal{E}\setminus H}1-p_e$. That is $H\in\mathcal{H}$ can be considered a random hypergraph with "individual" edge probabilities.

Furthermore, let $\mathcal{A}$ be a family of subsets of V, and let $(Y_\alpha)_{\alpha\in\mathcal{A}}$ be a family of random variables. Remember that for $Q\subset V$ we set $\mathcal{E}(Q)=\{e\in\mathcal{E}: e\cap Q\neq\emptyset\}$. We say that Y_α is *feasible* if the following holds.

For any two elements $H,H'\in\mathcal{H}$ such that $H\cap\mathcal{E}(\alpha)=H'\cap\mathcal{E}(\alpha)$ we have $Y_\alpha(H)=Y_\alpha(H')$.

That means Y_α is feasible if its value depends only on edges having at least one endpoint in α. In addition, set $Y_\alpha^S(H)=Y_\alpha(H\setminus\mathcal{E}(S))$ ($H\in\mathcal{H}$, $\alpha\in\mathcal{A}$, $S\subset V$, $S\cap\alpha=\emptyset$). Thus $Y_\alpha^S(H)$ is the value of Y_α after removing all edges incident with S. We define

$$Y=\sum_{\alpha\in\mathcal{A}}Y_\alpha,\quad \mu_\alpha=\mathrm{E}\left[Y_\alpha\right],\quad \sigma^2=\mathrm{Var}\left[Y\right],\quad X_\alpha=(Y_\alpha-\mu_\alpha)/\sigma \tag{54}$$

$$Z_\alpha=\sum_{\beta\in\mathcal{A}}Z_{\alpha\beta},\quad \text{where } Z_{\alpha\beta}=\sigma^{-1}\times\begin{cases}Y_\beta & \text{if } \alpha\cap\beta\neq\emptyset,\\ Y_\beta-Y_\beta^\alpha & \text{if } \alpha\cap\beta=\emptyset,\end{cases} \tag{55}$$

$$V_{\alpha\beta}=\sum_{\substack{\gamma:\beta\cap\gamma\neq\emptyset\\ \wedge\alpha\cap\gamma=\emptyset}}Y_\gamma^\alpha/\sigma+\sum_{\substack{\gamma:\beta\cap\gamma=\emptyset\\ \wedge\alpha\cap\gamma=\emptyset}}(Y_\gamma^\alpha-Y_\gamma^{\alpha\cup\beta})/\sigma,\qquad\text{and} \tag{56}$$

$$\delta=\sum_{\alpha\in\mathcal{A}}\mathrm{E}\left[|X_\alpha|\,Z_\alpha^2\right]+\sum_{\alpha,\beta\in\mathcal{A}}\left(\mathrm{E}\left[|X_\alpha Z_{\alpha\beta}V_{\alpha\beta}|\right]+\mathrm{E}\left[|X_\alpha Z_{\alpha\beta}|\right]\mathrm{E}\left[|Z_\alpha+V_{\alpha\beta}|\right]\right). \tag{57}$$

The following theorem was proven for graphs (i.e. $d=2$) in [3]. The argument used there carries over to the case of hypergraphs without essential modifications. Thus for the sake of brevity we omit a detailed proof of this result.

Theorem 26. *Suppose that all Y_α are feasible. If $\delta=o(1)$ as $n\to\infty$, then $\frac{Y-\mathrm{E}[Y]}{\sigma}$ converges to the standard normal distribution.*

Now the following lemma states that a number of conditions on the expectations of the product of up to three random variables Y_α^S will suffice for $\delta=o(1)$. The conditions are identical for both statements we want to prove and we will prove that they are fulfilled in both cases in the next two sections while the proof of the lemma itself is deferred to the end of the section.

Lemma 27. *Let $k=O(\mathrm{polylog}\,n)$ and let $(Y_\alpha)_{\alpha\in\mathcal{A}}$ be a feasible family such that $0\le Y_\alpha\le k$ for all $\alpha\in\mathcal{A}$. If the following six conditions are satisfied, then $\delta=o(1)$ as $n\to\infty$.*

Y1. *We have* $\mathrm{E}(Y),\mathrm{Var}(Y)=\Theta(n)$*, and* $\sum_{\beta\in\mathcal{A}:\beta\cap\alpha\neq\emptyset}\mu_\beta=O(\mathrm{E}(Y)/n\cdot\mathrm{polylog}\,n)=O(\mathrm{polylog}\,n)$*.for any* $\alpha\in\mathcal{A}$

Y2. *Let α,β,γ be distinct elements of $\mathcal{A}$. Then*

$$Y_\alpha(Y_\beta-Y_\beta^\alpha)Y_\beta^\alpha=0\qquad\textit{if } \alpha\cap\beta=\emptyset, \tag{58}$$

$$Y_\alpha Y_\beta=0\qquad\textit{if } \alpha\cap\beta\neq\emptyset, \tag{59}$$

$$(Y_\beta-Y_\beta^\alpha)Y_\gamma^\alpha=(Y_\beta-Y_\beta^\alpha)Y_\gamma=0\qquad\textit{if } \alpha\cap\beta=\alpha\cap\gamma=\emptyset\neq\beta\cap\gamma. \tag{60}$$

Y3. *For all* α, β *we have* $\sum_{\gamma:\gamma\cap\beta\neq\emptyset,\,\gamma\cap\alpha=\emptyset} \mathrm{E}(Y_\beta Y_\gamma^\alpha) \le k^2\mu_\beta$.
Y4. *If* $\alpha, \beta \in \mathcal{A}$ *are disjoint, then*

$$\mathrm{E}\left[Y_\alpha Y_\beta\right] = O(\mu_\alpha\mu_\beta \cdot \operatorname{polylog} n), \tag{61}$$
$$\mathrm{E}\left[|Y_\beta - Y_\beta^\alpha|\right] = O(\frac{\mu_\beta}{n} \cdot \operatorname{polylog} n), \tag{62}$$
$$\mathrm{E}\left[Y_\alpha|Y_\beta - Y_\beta^\alpha|\right] = O(\frac{\mu_\alpha\mu_\beta}{n} \cdot \operatorname{polylog} n). \tag{63}$$

Y5. *If* $\alpha, \beta, \gamma \in \mathcal{A}$ *are pairwise disjoint, then*

$$\mathrm{E}\left[Y_\beta|Y_\gamma^\alpha - Y_\gamma^{\alpha\cup\beta}|\right] = O(\frac{\mu_\beta\mu_\gamma}{n} \cdot \operatorname{polylog} n), \tag{64}$$
$$\mathrm{E}\left[|Y_\beta - Y_\beta^\alpha| \cdot |Y_\gamma^\alpha - Y_\gamma^{\alpha\cup\beta}|\right] = O(\frac{\mu_\beta\mu_\gamma}{n^2} \cdot \operatorname{polylog} n), \tag{65}$$
$$\mathrm{E}\left[Y_\alpha|Y_\beta - Y_\beta^\alpha| \cdot |Y_\gamma^\alpha - Y_\gamma^{\alpha\cup\beta}|\right] = O(\frac{\mu_\alpha\mu_\beta\mu_\gamma}{n^2} \cdot \operatorname{polylog} n), \tag{66}$$
$$\mathrm{E}\left[Y_\alpha|Y_\beta - Y_\beta^\alpha| \cdot |Y_\gamma - Y_\gamma^\alpha|\right] = O(\frac{\mu_\alpha\mu_\beta\mu_\gamma}{n^2} \cdot \operatorname{polylog} n), \tag{67}$$
$$\mathrm{E}\left[|(Y_\beta - Y_\beta^\alpha)(Y_\gamma - Y_\gamma^\alpha)|\right] = O(\frac{\mu_\alpha\mu_\beta}{n^2} \cdot \operatorname{polylog} n). \tag{68}$$

Y6. *If* $\alpha, \beta, \gamma \in \mathcal{A}$ *satisfy* $\alpha \cap \beta = \alpha \cap \gamma = \emptyset$, *then*

$$\mathrm{E}\left[|Y_\alpha^\beta - Y_\alpha^{\beta\cup\gamma}|\right] = O(\frac{\mu_\gamma}{n} \cdot \operatorname{polylog} n). \tag{69}$$

5.2 Conditions for the normality of $\mathcal{N}(H_d(n,p))$

In this section we will prove the properties **Y1–Y6** defined in Lemma 27 for the case of the normality of $\mathcal{N}(H_d(n,p))$.

Let $k = O(\operatorname{polylog} n)$ and let $\mathcal{A} = \{\alpha \subset V : 1 \le |\alpha| \le k\}$. Moreover, for $A \subseteq V$ with $A \cap \alpha = \emptyset$ let $I_\alpha^A = 1$ if α is a component of $H \setminus \mathcal{E}(A)$, and 0 otherwise. Further, set $Y_\alpha^A = |\alpha| \cdot I_\alpha^A$. We briefly write $I_\alpha = I_\alpha^\emptyset$ and $Y_\alpha = Y_\alpha^\emptyset$. Then $(Y_\alpha)_{\alpha\in\mathcal{A}}$ is a feasible family, because whether α is a component or not only depends on the presence of edges that contain at least one vertex of α.

Let $\mathcal{C}(S)$ denote the even that the subhypergraph of H induced on $S \subset V$ is connected. If $I_\alpha = 1$, then $\mathcal{C}(\alpha)$ occurs. Moreover, H contains no edges joining α and $V \setminus \alpha$, i.e., $H \cap \mathcal{E}(\alpha, V \setminus \alpha) = \emptyset$. Since each edge occurs in H with probability p independently, we thus obtain

$$\mathrm{P}\left[I_\alpha = 1\right] = \mathrm{P}\left[\mathcal{C}(\alpha)\right](1-p)^{|\mathcal{E}(\alpha, V\setminus\alpha)|}. \tag{70}$$

Furthermore, observe that

$$\forall \alpha \in \mathcal{A},\ A \subset B \subset V \setminus \alpha : I_\alpha^A = 1 \to I_\alpha^B = 1. \tag{71}$$

Proof of Y1: We know from Theorem 1 that $\mathrm{E}\left[Y\right] = \Theta(n)$ and $\mathrm{Var}\left[Y\right] = \Theta(n)$. To see that

$$\sum_{\beta\in\mathcal{A}:\beta\cap\alpha\neq\emptyset} \mu_\beta = O(\mathrm{E}\left[Y\right]/n \cdot \operatorname{polylog} n),$$

note that $\mu_\beta := \mathrm{E}\left[Y_\beta\right]$ depends only on the size of β. Thus with $\mu_b = \mu_\beta$ for an arbitrary set β of size b we have $\mathrm{E}\left[Y\right] = \sum_{\beta\in\mathcal{A}} \mu_\beta = \sum_{b=1}^k \sum_{\substack{\beta\in\mathcal{A} \\ |\beta|=b}} \mu_\beta = \sum_{b=1}^k \binom{n}{b}\mu_b$ while $\sum_{\beta\in\mathcal{A}:\beta\cap\alpha\neq\emptyset} \mu_\beta = \sum_{b=1}^k \sum_{\substack{\beta\cap\alpha\neq\emptyset \\ |\beta|=b}} \mu_\beta \le \sum_{b=1}^k k\binom{n}{b-1}\mu_b$.

Proof of Y2: (58): Suppose that $I_\alpha = 1$. Then H features no edge that contains a vertex in α and a vertex in β. If in addition $I_\beta^\alpha = 1$, then we obtain that $I_\beta = 1$ as well. Hence, $Y_\beta = Y_\beta^\alpha$.

(59): This just means that any two components of H are either disjoint or equal.

(60): To show that $Y_\gamma(Y_\beta - Y_\beta^\alpha) = 0$, assume that $I_\gamma = 1$. Then γ is a component of H, so that β cannot be a component, because $\gamma \neq \beta$ but $\gamma \cap \beta \neq \emptyset$; hence, $I_\beta = 0$. Furthermore, if γ is a component of H, then γ is also a component of $H \setminus \mathcal{E}(\alpha)$, so that $I_\gamma^\alpha = 1$. Consequently, $I_\beta^\alpha = 0$. Thus, $Y_\beta = Y_\beta^\alpha = 0$.

In order to prove that $Y_\gamma^\alpha(Y_\beta - Y_\beta^\alpha) = 0$, suppose that $I_\gamma^\alpha = 1$. Then $I_\beta^\alpha = 0$, because the intersecting sets β, γ cannot both be components of $H \setminus \mathcal{E}(\alpha)$. Therefore, we also have $I_\beta = 0$; for if β were a component of H, then β would also be a component of $H \setminus \mathcal{E}(\alpha)$. Hence, also in this case we obtain $Y_\beta = Y_\beta^\alpha = 0$.

Proof of Y3: Suppose that $I_\beta = 1$, i.e., β is a component of H. Then removing the edges $\mathcal{E}(\alpha)$ from H may cause β to split into several components $B_1, \ldots, B_l$. Thus, if $Y_\gamma^\beta > 0$ for some $\gamma \in \mathcal{A}$ such that $\gamma \cap \beta \neq \emptyset$, then γ is one of the components $B_1, \ldots, B_l$. Since $l \leq |\beta| \leq k$, this implies that given $I_\beta = 1$ we have the bound $\sum_{\gamma:\gamma\cap\beta\neq\emptyset,\,\gamma\cap\alpha=\emptyset} Y_\gamma^\alpha \leq k^2$. Hence, we obtain **Y3**.

The following lemma which gives a description of the limited dependence between the random variables I_α and I_β for disjoint α and β together with the fact that $\mathrm{P}[I_\alpha = 1] = O(\mu_\alpha)$ implies **Y4–Y6**.

Lemma 28. *Let $0 \leq l, r \leq 2$, and let $\alpha_1, \ldots, \alpha_l, \beta_1, \ldots, \beta_r \in \mathcal{A}$ be pairwise disjoint. Moreover, let $A_1, \ldots, A_r, B_1, \ldots, B_r \subset V$ be sets such that $A_i \subset B_i \subset V \setminus \beta_i$ and $|B_i| \leq 2k$ for all $1 \leq i \leq r$, and assume that $\bigcap_{i=1}^r B_i \setminus A_i = \emptyset$. Then*

$$\mathrm{P}\left[\bigwedge_{i=1}^{l}, \bigwedge_{j=1}^{r} I_{\alpha_i} = 1 \wedge I_{\beta_j}^{A_j} \neq I_{\beta_j}^{B_j}\right] \leq O(n^{-r} \cdot \mathrm{polylog}\, n) \prod_{j=1}^{l} \mathrm{P}\left[I_{\alpha_i} = 1\right] \prod_{j=1}^{r} \mathrm{P}\left[I_{\beta_j} = 1\right].$$

Proof. Since (71) entails that $I_{\beta_j}^{A_j} \neq I_{\beta_j}^{B_j} \leftrightarrow I_{\beta_j}^{B_j} = 1 \wedge I_{\beta_j}^{A_j} = 0$, we have

$$\mathrm{P}\left[\forall i, j : I_{\alpha_i} = 1 \wedge I_{\beta_j}^{A_j} \neq I_{\beta_j}^{B_j}\right] = \mathrm{P}\left[\forall i, j : I_{\alpha_i} = 1 \wedge I_{\beta_j}^{A_j} = 0 \wedge I_{\beta_j}^{B_j} = 1\right]. \tag{72}$$

We shall bound the probability on the right hand side in terms of mutually independent events.

If $I_{\alpha_i} = 1$ and $I_{\beta_j}^{B_j} = 1$ for all i, j, then the hypergraphs induced on α_i and β_j are connected, i.e., the events $\mathcal{C}(\alpha_i)$ and $\mathcal{C}(\beta_j)$ occur. Note that these events are mutually independent, because $\mathcal{C}(\alpha_i)$ (resp. $\mathcal{C}(\beta_j)$) *only* depends on the presence of edges $e \in \mathcal{E}(\alpha_i) \setminus \mathcal{E}(V \setminus \alpha_i)$ (resp. $e \subset \mathcal{E}(\beta_j) \setminus \mathcal{E}(V \setminus \beta_j)$).

Furthermore, if α_i is a component, then in H there occur no edges joining α_i and $V \setminus \alpha_i$; in other words, $H \cap \mathcal{E}(\alpha_i, V \setminus \alpha_i) = \emptyset$. However, these events are not necessarily independent, because $\mathcal{E}(\alpha_1, V \setminus \alpha_1)$ may contain edges that are incident with vertices in α_2. Therefore, we consider the sets

$$\mathcal{F}(\alpha_i) = \bigcup_{i' \neq i} \alpha_{i'} \cup \bigcup_{j=1}^{r} \beta_j \cup B_j,\ \ \mathcal{D}(\alpha_i) = \mathcal{E}(\alpha_i, V \setminus \alpha_i) \setminus \mathcal{E}(\mathcal{F}(\alpha_i)),$$
$$\mathcal{F}(\beta_j) = \bigcup_{i=1}^{l} \alpha_i \cup \bigcup_{j' \neq j} \beta_{j'} \cup \bigcup_{j'=1}^{r} B_{j'},\ \ \mathcal{D}(\beta_j) = \mathcal{E}(\beta_j, V \setminus \beta_j) \setminus \mathcal{E}(\mathcal{F}(\beta_j)).$$

Then $I_{\alpha_i} = 1$ (resp. $I_{\beta_j}^{B_j} = 1$) implies that $\mathcal{D}(\alpha_i) \cap H = \emptyset$ (resp. $\mathcal{D}(\beta_j) \cap H = \emptyset$). Moreover, since the sets $\mathcal{D}(\alpha_i)$ and $\mathcal{D}(\beta_j)$ are pairwise disjoint, the events $\mathcal{D}(\alpha_i) \cap H = \emptyset$, $\mathcal{D}(\beta_i) \cap H = \emptyset$ are mutually independent.

Finally, we need to express the fact that $I_{\beta_j}^{A_j} = 0$ but $I_{\beta_j}^{B_j} = 1$. If this event occurs, then H contains an edge connecting β_j with $B_j \setminus A_j$, i.e., $H \cap \mathcal{E}(\beta_j, B_j \setminus A_j) \neq \emptyset$. Thus, let $\mathcal{Q}$ denote the event that $H \cap \mathcal{E}(\beta_j, B_j \setminus A_j) \neq \emptyset$ for all $1 \leq j \leq r$.

Thus, we obtain

$$\begin{aligned}
&\mathrm{P}\left[\forall i, j : I_{\alpha_i} = 1 \wedge I_{\beta_j}^{A_j} = 0 \wedge I_{\beta_j}^{B_j} = 1\right] \\
\leq\ &\mathrm{P}\left[\bigwedge_{i=1}^{l} (\mathcal{C}(\alpha_i) \wedge (\mathcal{D}(\alpha_i) \cap H = \emptyset)) \wedge \bigwedge_{j=1}^{r} (\mathcal{C}(\beta_j) \wedge (\mathcal{D}(\beta_j) \cap H = \emptyset)) \wedge \mathcal{Q}\right] \\
=\ &\prod_{i=1}^{l} \mathrm{P}\left[\mathcal{C}(\alpha_i)\right] \mathrm{P}\left[\mathcal{D}(\alpha_i) \cap H = \emptyset\right] \times \prod_{j=1}^{r} \mathrm{P}\left[\mathcal{C}(\beta_j)\right] \mathrm{P}\left[\mathcal{D}(\beta_j) \cap H = \emptyset\right] \times \mathrm{P}\left[\mathcal{Q}\right].
\end{aligned} \tag{73}$$

We shall prove below that

$$\mathrm{P}\left[\mathcal{D}(\alpha_i)\cap H=\emptyset\right]\sim(1-p)^{|\mathcal{E}(\alpha_i,V\setminus\alpha_i)|},\quad \mathrm{P}\left[\mathcal{D}(\beta_j)\cap H=\emptyset\right]\sim(1-p)^{|\mathcal{E}(\beta_j,V\setminus\beta_j)|}, \tag{74}$$
$$\mathrm{P}\left[\mathcal{Q}\right]=O(n^{-r}\cdot\mathrm{polylog}\,n). \tag{75}$$

Combining (70) and (72)–(75), we then obtain the assertion.

To establish (74), note that by definition $\mathcal{D}(\alpha_i)\subset\mathcal{E}(\alpha_i,V\setminus\alpha_i)$. Therefore,

$$\mathrm{P}\left[\mathcal{D}(\alpha_i)\cap H=\emptyset\right]=(1-p)^{|\mathcal{D}(\alpha_i)|}\geq(1-p)^{|\mathcal{E}(\alpha_i,V\setminus\alpha_i)|}. \tag{76}$$

On the other hand, we have $|\alpha_i|,|\mathcal{F}(\alpha_i)|=O(\mathrm{polylog}\,n)$, and thus $|\mathcal{E}(\alpha_i,\mathcal{F}(\alpha_i))|\leq|\alpha_i|\cdot|\mathcal{F}(\alpha_i)|\cdot\binom{n}{d-2}=O(n^{d-2}\cdot\mathrm{polylog}\,n)$. Hence, as $p=O(n^{1-d})$, we obtain

$$\begin{aligned}\mathrm{P}\left[\mathcal{D}(\alpha_i)\cap H=\emptyset\right]=(1-p)^{|\mathcal{D}(\alpha_i)|}&\leq(1-p)^{|\mathcal{E}(\alpha_i,V\setminus\alpha_i)|-|\mathcal{E}(\alpha_i,\mathcal{F}(\alpha_i))|}\\ &\sim(1-p)^{|\mathcal{E}(\alpha_i,V\setminus\alpha_i)|}\exp(p\cdot O(n^{d-2}\cdot\mathrm{polylog}\,n))\sim(1-p)^{|\mathcal{E}(\alpha_i,V\setminus\alpha_i)|}.\end{aligned} \tag{77}$$

Combining (76) and (77), we conclude that $\mathrm{P}\left[\mathcal{D}(\alpha_i)\cap H=\emptyset\right]\sim(1-p)^{|\mathcal{E}(\alpha_i,V\setminus\alpha_i)|}$. As the same argument applies to $\mathrm{P}\left[\mathcal{D}(\beta_j)\cap H=\emptyset\right]$, we thus obtain (74).

Finally, we prove (75). If $r=1$, then H contains an edge of $\mathcal{E}(\beta_1,B_1\setminus A_1)$. Since

$$|\mathcal{E}(\beta_1,B_1\setminus A_1)|\leq|\beta_1|\cdot|B_1\setminus A_1|\cdot n^{d-2}=O(n^{d-2}\cdot\mathrm{polylog}\,n),$$

and because each possible edge occurs with probability p independently, the probability of this event is $\mathrm{P}\left[\mathcal{Q}\right]\leq O(n^{d-2}\cdot\mathrm{polylog}\,n)p=O(n^{-1}\cdot\mathrm{polylog}\,n)$, as desired.

Now, assume that $r=2$. Then H features edges $e_j\in\mathcal{E}(\beta_j,B_j\setminus A_j)$ $(j=1,2)$.

1st case: $e_1=e_2$**.** In this case, e_1 contains a vertex of each of the four sets β_1, β_2, $B_1\setminus A_1$, $B_2\setminus A_2$. Hence, the number of possible such edges is $\leq n^{d-4}\prod_{j=1}^{2}|\beta_j|\cdot|B_j\setminus A_j|=O(n^{d-4}\cdot\mathrm{polylog}\,n)$. Consequently, the probability that such an edge occurs in H is $\leq O(n^{d-4}\cdot\mathrm{polylog}\,n)p=O(n^{-3}\cdot\mathrm{polylog}\,n)$.

2nd case: $e_1\neq e_2$**.** There are $\leq|\beta_j|\cdot|B_j\setminus A_j|\cdot n^{d-2}=O(n^{d-2}\cdot\mathrm{polylog}\,n)$ ways to choose e_j $(j=1,2)$. Hence, the probability that such edges e_1,e_2 occur in H is $\leq\left[O(n^{d-2}\cdot\mathrm{polylog}\,n)p\right]^2=O(n^{-2}\cdot\mathrm{polylog}\,n)$.

Thus, in both cases we obtain the bound claimed in (75). □

5.3 Conditions for the normality of $\mathcal{S}_G$

In this section we will prove the properties **Y1–Y6** defined in Lemma 27 for the case of the normality of $\mathcal{S}_G$.

Consider a set $G\subset V$ of size n_1. Let $\mathcal{A}$ be the set of all subsets $\alpha\subset V\setminus G$ of size $|\alpha|\leq k$. Moreover, let $p_e=p$ for $e\subset V\setminus G$, $p_e=p_2$ for $e\in\mathcal{E}(G,V\setminus G)$, and $p_e=0$ if $e\subset G$.

For $A\subseteq V$ and $A\cap\alpha=\emptyset$ set $I_\alpha^A=1$ if α is a component of $H\setminus\mathcal{E}(A\cup G)$. Moreover, let $J_\alpha^A=1$ if $(H\setminus\mathcal{E}(A))\cap\mathcal{E}(G,\alpha)\neq\emptyset$. Further, let $K_\alpha^A=I_\alpha^A J_\alpha^A$ and $Y_\alpha^A=|\alpha|K_\alpha^A$. Then

$$\mathrm{P}\left[K_\alpha=1\right]=\Omega(\mathrm{P}\left[I_\alpha=1\right]). \tag{78}$$

Proof of Y1: Using Lemma 10 we have $\mathrm{E}\left[Y\right]=\Theta(n)$ and using Lemma 11 we have $\mathrm{Var}\left[Y\right]=\Theta(n)$. The proof of the rest of **Y1** is analogous to the proof of **Y1** in the case of $\mathcal{N}(H_d(n,p))$.

Proof of Y2: (58): Suppose that $K_\alpha=1$. Then $I_\alpha=1$, so that $H\setminus\mathcal{E}(G)$ has no α-β-edges. Hence, if also $K_\beta^\alpha=1$, then β is a component of $H\setminus\mathcal{E}(G)$ as well. Thus, $K_\beta=1$, so that $Y_\beta=Y_\beta^\alpha$.

(59): If $K_\alpha=1$, then α is a component of $H\setminus\mathcal{E}(G)$. Since any two components of $H\setminus\mathcal{E}(G)$ are either disjoint or equal, we obtain $I_\beta=0$, so that $Y_\beta=0$ as well.

(60): To show that $Y_\gamma(Y_\beta-Y_\beta^\alpha)=0$, assume that $K_\gamma=1$. Then $I_\gamma=1$, i.e., γ is a component of $H\setminus\mathcal{E}(G)$. Since $\beta\neq\gamma$ but $\beta\cap\gamma\neq\emptyset$, we conclude that $I_\beta=0$. Furthermore, if γ is a component of $H\setminus\mathcal{E}(G)$, then γ is also a component of $H\setminus\mathcal{E}(G\cup\alpha)$, whence $I_\beta^\alpha=0$. Consequently, $Y_\beta=Y_\beta^\alpha=0$.

In order to prove that $Y_\gamma^\alpha(Y_\beta - Y_\beta^\alpha) = 0$, suppose that $K_\gamma^\alpha = 1$. Then $K_\gamma^\alpha = 1$. Therefore, $I_\beta^\alpha = 0$, because the intersecting sets β, γ cannot both be components of $H \setminus \mathcal{E}(\alpha)$. Thus, we also have $I_\beta = 0$; for if β were a component of H, then β would also be a component of $H \setminus \mathcal{E}(\alpha)$. Hence, also in this case we obtain $Y_\beta = Y_\beta^\alpha = 0$.

Proof of Y3: Suppose that $K_\beta = 1$. Then $I_\beta = 1$, i.e., β is a component of $H \setminus \mathcal{E}(G)$. Then removing the edges $\mathcal{E}_\alpha$ from $H \setminus \mathcal{E}(G)$ may cause β to split into several components $B_1, \ldots, B_l$. Thus, if $Y_\gamma^\beta > 0$ for some $\gamma \in \mathcal{A}$ such that $\gamma \cap \beta \neq \emptyset$, then γ is one of the components $B_1, \ldots, B_l$. Since $l \leq |\beta| \leq k$, this implies that given $I_\beta = 1$ we have the bound

$$\sum_{\gamma : \gamma \cap \beta \neq \emptyset,\, \gamma \cap \alpha = \emptyset} Y_\gamma^\alpha \leq k^2.$$

Hence, we obtain **Y3**.

Similar to Lemma 28 the following lemma on the limited dependence of K_α and K_β for disjoint α and β implies **Y4–Y6**.

Lemma 29. *Let $0 \leq l, r \leq 2$, and let $\alpha_1, \ldots, \alpha_l, \beta_1, \ldots, \beta_r \in \mathcal{A}$ be pairwise disjoint. Moreover, let $A_1, \ldots, A_r, B_1, \ldots, B_r \subset V$ be sets such that $A_i \subset B_i \subset V \setminus \beta_i$ and $|B_i| \leq O(1)$ for all $1 \leq i \leq r$, and assume that $\bigcap_{i=1}^r B_i \setminus A_i = \emptyset$. Then*

$$\mathrm{P}\left[\bigwedge_{i=1}^{l} \bigwedge_{j=1}^{r} K_{\alpha_i} = 1 \wedge K_{\beta_j}^{A_j} \neq K_{\beta_j}^{B_j}\right] \leq O(n^{-r} \cdot \mathrm{polylog}\, n) \prod_{j=1}^{l} \mathrm{P}\left[K_{\alpha_i} = 1\right] \prod_{j=1}^{r} \mathrm{P}\left[K_{\beta_j} = 1\right].$$

Proof. Let $\mathcal{P} = \mathrm{P}\left[\forall i, j : K_{\alpha_i} = 1 \wedge K_{\beta_j}^{A_j} \neq K_{\beta_j}^{B_j}\right]$. If $K_{\beta_j}^{A_j} \neq K_{\beta_j}^{B_j}$, then either $I_{\beta_j}^{A_j} \neq I_{\beta_j}^{B_j}$ or $I_{\beta_j}^{A_j} = I_{\beta_j}^{B_j} = 1$ and $J_{\beta_j}^{A_j} \neq J_{\beta_j}^{B_j}$. Therefore, letting $\mathcal{J} = \{j : I_{\beta_j}^{A_j} \neq I_{\beta_j}^{B_j}\}$ and $\bar{\mathcal{J}} = \{1, \ldots, r\} \setminus \mathcal{J}$, we obtain

$$\mathcal{P} \leq \mathrm{P}\left[\bigwedge_{i=1}^{l} I_{\alpha_i} = 1 \wedge \bigwedge_{j \in \mathcal{J}} I_{\beta_j}^{A_j} \neq I_{\beta_j}^{B_j} \wedge \bigwedge_{j \in \bar{\mathcal{J}}} \left(I_{\beta_j}^{A_j} = 1 \wedge J_{\beta_j}^{A_j} \neq J_{\beta_j}^{B_j}\right)\right]. \tag{79}$$

Now, the random variables I_{α_i}, $I_{\beta_j}^{A_j}$, and $I_{\beta_j}^{B_j}$ are determined just by the edges in $\mathcal{E} \setminus \mathcal{E}(G)$, while $J_{\beta_j}^{A_j}$ and $J_{\beta_j}^{B_j}$ depend only on the edges in $\mathcal{E}(G)$. Hence, as the edges in $\mathcal{E} \setminus \mathcal{E}(G)$ and in $\mathcal{E}(G)$ occur in H independently, (79) yields

$$\mathcal{P} \leq \mathrm{P}\left[\bigwedge_{i=1}^{l} I_{\alpha_i} = 1 \wedge \bigwedge_{j \in \bar{\mathcal{J}}} I_{\beta_j}^{A_j} = 1 \wedge \bigwedge_{j \in \mathcal{J}} I_{\beta_j}^{A_j} \neq I_{\beta_j}^{B_j}\right] \cdot \mathrm{P}\left[\bigwedge_{j \in \bar{\mathcal{J}}} J_{\beta_j}^{A_j} \neq J_{\beta_j}^{B_j}\right]. \tag{80}$$

Furthermore, Lemma 28 entails that

$$\mathrm{P}\left[\bigwedge_{i=1}^{l} I_{\alpha_i} = 1 \wedge \bigwedge_{j \in \bar{\mathcal{J}}} I_{\beta_j}^{A_j} = 1 \wedge \bigwedge_{j \in \mathcal{J}} I_{\beta_j}^{A_j} \neq I_{\beta_j}^{B_j}\right] \leq O(n^{-|\mathcal{J}|} \cdot \mathrm{polylog}\, n) \cdot \prod_{i=1}^{l} \mathrm{P}\left[I_{\alpha_i} = 1\right] \cdot \prod_{j=1}^{r} \mathrm{P}\left[I_{\beta_j} = 1\right]. \tag{81}$$

In addition, we shall prove below that

$$\mathrm{P}\left[\bigwedge_{j \in \bar{\mathcal{J}}} J_{\beta_j}^{A_j} \neq J_{\beta_j}^{B_j}\right] \leq O(n^{-|\bar{\mathcal{J}}|} \cdot \mathrm{polylog}\, n). \tag{82}$$

Plugging (81) and (82) into (80), we get $\mathcal{P} \leq O(n^{-r} \cdot \mathrm{polylog}\, n) \cdot \prod_{i=1}^{l} \mathrm{P}\left[I_{\alpha_i} = 1\right] \cdot \prod_{j=1}^{r} \mathrm{P}\left[I_{\beta_j} = 1\right]$, so that the assertion follows from (78).

Thus, the remaining task is to establish (82). Let us first deal with the case $|\bar{\mathcal{J}}| = 1$. Let $j \in \bar{\mathcal{J}}$. If $J_{\beta_j}^{A_j} \neq J_{\beta_j}^{B_j}$, then $J_{\beta_j}^{A_j} = 1$ and $J_{\beta_j}^{B_j} = 0$, because $A_j \subset B_j$. Thus, β_j is connected to G via an

edge that is incident with $A_j \setminus B_j$; that is, $H \cap \mathcal{E}(\beta_j, B_j \setminus A_j) \neq \emptyset$. Since there are $|\mathcal{E}(\beta_j, B_j \setminus A_j)| \leq |\beta_j| \cdot |B_j| \cdot n^{d-2} = O(n^{d-2} \cdot \operatorname{polylog} n)$ such edges to choose from, and because each such edge is present with probability $p_2 = O(n^{1-d})$, we conclude that $\mathrm{P}\left[J_{\beta_j}^{A_j} \neq J_{\beta_j}^{B_j}\right] \leq \mathrm{P}\left[H \cap \mathcal{E}(\beta_j, B_j \setminus A_j) \neq \emptyset\right] \leq O(n^{d-2} \cdot \operatorname{polylog} n) p_2 = O(n^{-1} \cdot \operatorname{polylog} n)$, whence we obtain (82).

Finally, suppose that $|\bar{\mathcal{J}}| = 2$. If $J_{\beta_j}^{A_j} \neq J_{\beta_j}^{B_j}$ for $j = 1, 2$, then there occur edges $e_j \in H \cap \mathcal{E}(\beta_j, B_j \setminus A_j)$ $(j = 1, 2)$.

1st case: $e_1 = e_2$**.** In this case $e_1 = e_2$ is incident with all four sets $\beta_j, B_j \setminus A_j$ $(j = 1, 2)$. Hence, as the number of such edges is $\leq n^{d-4} \prod_{j=1}^{2} |\beta_j| \cdot |B_j \setminus A_j| \leq O(n^{d-4} \cdot \operatorname{polylog} n)$ and each such edge occurs with probability $p_2 = O(n^{1-d})$, the probability that the 1st case occurs is $O(n^{d-4} \cdot \operatorname{polylog} n) p_2 = O(n^{-3} \cdot \operatorname{polylog} n)$.

2nd case: $e_1 \neq e_2$**.** There are $\leq |\beta_j| \cdot |B_j \setminus A_j| \cdot n^{d-2} \leq O(n^{d-2} \cdot \operatorname{polylog} n)$ ways to choose e_j for $j = 1, 2$, each of which is present with probability $p_2 = O(n^{1-d})$ independently. Hence, the probability that the second case occurs is bounded by $\left[O(n^{d-2} \cdot \operatorname{polylog} n) p_2\right]^2 \leq O(n^{-2} \cdot \operatorname{polylog} n)$.

Thus, the bound (82) holds in both cases. $\square$

5.4 Proof of Lemma 27

All we need to show is that the conditions defined in Lemma 27 imply that δ as defined by (57) tends to 0. We will do so by proving that each of the three summands contributing to δ is $O(\sigma^{-3} \mathrm{E}[Y] \cdot \operatorname{polylog} n)$. Together with condition **Y1**, stating that $\mathrm{E}[Y], \sigma^2 = \Theta(n)$, this implies the statement. We formulate one lemma for each summand, bounding the expectations using conditions **Y1–Y6**. The proof of the lemmas are mainly long and technical computations then.

Lemma 30. $\sum_{\alpha \in \mathcal{A}} \mathrm{E}\left[|X_\alpha| Z_\alpha^2\right] = O(\sigma^{-3} \mathrm{E}[Y] \cdot \operatorname{polylog} n)$

Proof. Let

$$S_1 = \sum_{\alpha \in \mathcal{A}} \mathrm{E}\left[Y_\alpha \left(\sum_{\beta : \alpha \cap \beta \neq \emptyset} Y_\beta\right)^2\right], \; S_2 = \sum_{\alpha \in \mathcal{A}} \mathrm{E}\left[\mu_\alpha \left(\sum_{\beta : \alpha \cap \beta \neq \emptyset} Y_\beta\right)^2\right],$$

$$S_3 = \sum_{\alpha \in \mathcal{A}} \mathrm{E}\left[Y_\alpha \left(\sum_{\beta : \alpha \cap \beta = \emptyset} (Y_\beta - Y_\beta^\alpha)\right)^2\right], \; S_4 = \sum_{\alpha \in \mathcal{A}} \mathrm{E}\left[\mu_\alpha \left(\sum_{\beta : \alpha \cap \beta = \emptyset} (Y_\beta - Y_\beta^\alpha)\right)^2\right].$$

Since $X_\alpha = (Y_\alpha - \mu_\alpha)/\sigma \leq (Y_\alpha + \mu_\alpha)/\sigma$, (54) entails that

$$\begin{aligned} \mathrm{E}\left[|X_\alpha| Z_\alpha^2\right] &\leq 2\sigma^{-3} \mathrm{E}\left[(Y_\alpha + \mu_\alpha)\left(\left(\sum_{\beta : \alpha \cap \beta \neq \emptyset} Y_\beta\right)^2 + \left(\sum_{\beta : \alpha \cap \beta = \emptyset} (Y_\beta - Y_\beta^\alpha)\right)^2\right)\right] \\ &\leq 2\sigma^{-3}(S_1 + S_2 + S_3 + S_4). \end{aligned}$$

Therefore, it suffices to show that $S_j = O(\mathrm{E}(Y) \cdot \operatorname{polylog} n)$ for $j = 1, 2, 3, 4$.

Regarding S_1, we obtain the bound

$$S_1 = \sum_{\alpha \in \mathcal{A}} \sum_{\beta : \alpha \cap \beta \neq \emptyset} \sum_{\gamma : \alpha \cap \gamma \neq \emptyset} \mathrm{E}[Y_\alpha Y_\beta Y_\gamma] \overset{(59)}{\leq} k^2 \sum_{\alpha \in \mathcal{A}} \mathrm{E}[Y_\alpha] \leq O(\mathrm{E}[Y] \cdot \operatorname{polylog} n).$$

With respect to S_2, note that due to (59) and (61) we have $\mathrm{E}[Y_\beta Y_\gamma] \leq k\mu_\beta$ if $\beta = \gamma$, $\mathrm{E}[Y_\beta Y_\gamma] = 0$ if $\beta \neq \gamma$ but $\beta \cap \gamma \neq \emptyset$, and $\mathrm{E}[Y_\beta Y_\gamma] = O(\mu_\beta \mu_\gamma \cdot \operatorname{polylog} n)$ if $\beta \cap \gamma = \emptyset$. Consequently,

$$\begin{aligned} S_2 &= \sum_{\alpha \in \mathcal{A}} \mu_\alpha \sum_{\beta : \alpha \cap \beta \neq \emptyset} \sum_{\gamma : \alpha \cap \gamma \neq \emptyset} \mathrm{E}[Y_\beta Y_\gamma] \\ &\leq \sum_{\alpha \in \mathcal{A}} \mu_\alpha \sum_{\beta : \alpha \cap \beta \neq \emptyset} \sum_{\gamma : \alpha \cap \gamma \neq \emptyset} O(\mu_\beta \mu_\gamma \cdot \operatorname{polylog} n) \overset{\mathbf{Y1}}{\leq} O(\mathrm{E}(Y) \cdot \operatorname{polylog} n). \end{aligned} \tag{83}$$

Concerning S_3, we obtain

$$\begin{aligned} S_3 &= \sum_{\alpha\in\mathcal{A}} \sum_{\beta:\alpha\cap\beta=\emptyset} \sum_{\gamma:\alpha\cap\gamma=\emptyset} \mathrm{E}\left[Y_\alpha(Y_\beta - Y_\beta^\alpha)(Y_\gamma - Y_\gamma^\alpha)\right] \\ &\overset{(67),(60)}{\leq} \sum_{\alpha\in\mathcal{A}} \sum_{\beta:\alpha\cap\beta=\emptyset} \sum_{\gamma:\alpha\cap\gamma=\emptyset} O(\mu_\alpha\mu_\beta\mu_\gamma n^{-2}\cdot \mathrm{polylog}\, n) \\ &\leq O(n^{-2}\cdot \mathrm{polylog}\, n)\mathrm{E}(Y)^3 \overset{\mathbf{Y1}}{\leq} O(\mathrm{E}(Y)\cdot \mathrm{polylog}\, n). \end{aligned}$$

To bound S_4, we note that for disjoint $\alpha,\beta\in\mathcal{A}$ and $\gamma\in\mathcal{A}$ disjoint from α the conditions (62), (59), and (68) yield

$$\mathrm{E}\left[|(Y_\beta - Y_\beta^\alpha)(Y_\gamma - Y_\gamma^\alpha)|\right] = \begin{cases} O(\frac{\mu_\beta}{n}\cdot \mathrm{polylog}\, n) & \text{if } \beta=\gamma \\ 0 & \text{if } \beta\neq\gamma, \beta\cap\gamma\neq\emptyset \\ O(\frac{\mu_\beta\mu_\gamma}{n^2}\cdot \mathrm{polylog}\, n) & \text{if } \beta\cap\gamma=\emptyset. \end{cases}$$

Therefore,

$$\begin{aligned} \sum_{\beta:\alpha\cap\beta=\emptyset} \sum_{\gamma:\alpha\cap\gamma=\emptyset} \mathrm{E}\left[|(Y_\beta - Y_\beta^\alpha)(Y_\gamma - Y_\gamma^\alpha)|\right] &\leq \sum_{\beta\in\mathcal{A}}\sum_{\gamma\in\mathcal{A}} O(\frac{\mu_\beta\mu_\gamma}{n^2}\cdot \mathrm{polylog}\, n) + \sum_{\beta\in\mathcal{A}} O(\frac{\mu_\beta}{n}\cdot \mathrm{polylog}\, n) \\ &\leq O(\mathrm{E}(Y)^2/n^2\cdot \mathrm{polylog}\, n) + O(\mathrm{E}(Y)/n\cdot \mathrm{polylog}\, n) \\ &= O(\mathrm{polylog}\, n). \end{aligned}$$

Hence, we obtain $S_4 \leq \sum_{\alpha\in\mathcal{A}}\mu_\alpha \sum_{\beta:\alpha\cap\beta\neq\emptyset}\sum_{\gamma:\alpha\cap\gamma\neq\emptyset} \mathrm{E}\left[(Y_\beta - Y_\beta^\alpha)(Y_\gamma - Y_\gamma^\alpha)\right] \leq O(\mathrm{E}[Y]\cdot \mathrm{polylog}\, n)$. □

Lemma 31. $\sum_{\alpha\in\mathcal{A}}\sum_{\beta\in\mathcal{A}} \mathrm{E}\left[|X_\alpha Z_{\alpha\beta} V_{\alpha\beta}|\right] = O(\sigma^{-3}\mathrm{E}[Y]\cdot \mathrm{polylog}\, n)$

Proof. Let $S_1 = \sum_{\beta:\alpha\cap\beta\neq\emptyset}\mathrm{E}\left[|X_\alpha Y_\beta V_{\alpha\beta}|\right]$ and $S_2 = \sum_{\beta:\alpha\cap\beta=\emptyset}\mathrm{E}\left[\left|X_\alpha(Y_\beta - Y_\beta^\alpha)V_{\alpha\beta}\right|\right]$. Then the definition (55) of $Z_{\alpha\beta}$ yields that $\sum_{\alpha\in\mathcal{A}}\sum_{\beta\in\mathcal{A}}\mathrm{E}\left[|X_\alpha Z_{\alpha\beta}V_{\alpha\beta}|\right] \leq \sigma^{-1}(S_1+S_2)$ Hence, it suffices to show that $S_1, S_2 = O(\sigma^{-2}\mathrm{E}[Y]\cdot \mathrm{polylog}\, n)$.

To bound S_1, we note that $Y_\alpha Y_\beta = 0$ if $\alpha\cap\beta\neq\emptyset$ but $\alpha\neq\beta$ by (59), and that $V_{\alpha\beta}=0$ if $\alpha=\beta$ by the definition (56) of $V_{\alpha\beta}$. Thus, if $\alpha\cap\beta\neq\emptyset$, then

$$\mathrm{E}\left[|X_\alpha Y_\beta V_{\alpha\beta}|\right] \overset{(54)}{\leq} \sigma^{-1}\mathrm{E}\left[(Y_\alpha+\mu_\alpha)\,|Y_\beta V_{\alpha\beta}|\right] \leq \sigma^{-1}\mu_\alpha \mathrm{E}\left[|Y_\beta V_{\alpha\beta}|\right]. \tag{84}$$

Furthermore,

$$T_1(\alpha,\beta) = \sum_{\gamma:\gamma\cap\beta\neq\emptyset,\,\gamma\cap\alpha=\emptyset} \mathrm{E}\left[|Y_\beta Y_\gamma^\alpha\right] \overset{\mathrm{Y7}}{\leq} k^2\mu_\beta. \tag{85}$$

$$\begin{aligned} T_2(\alpha) = \sum_{\beta:\alpha\cap\beta\neq\emptyset} \sum_{\substack{\gamma:\beta\cap\gamma=\emptyset \\ \wedge\alpha\cap\gamma=\emptyset}} \mathrm{E}\left[Y_\beta|Y_\gamma^\alpha - Y_\gamma^{\alpha\cup\beta}|\right] &\overset{(64)}{\leq} \sum_{\beta:\alpha\cap\beta\neq\emptyset} \sum_{\substack{\gamma:\beta\cap\gamma=\emptyset \\ \wedge\alpha\cap\gamma=\emptyset}} O(\frac{\mu_\beta\mu_\gamma}{n}\cdot \mathrm{polylog}\, n) \\ &\leq O(n^{-1}\cdot \mathrm{polylog}\, n)\left[\sum_{\gamma\in\mathcal{A}}\mu_\gamma\right]\sum_{\beta:\alpha\cap\beta\neq\emptyset}\mu_\beta \\ &\overset{\mathbf{Y1}}{\leq} O(n^{-1}\mathrm{E}(Y)\cdot \mathrm{polylog}\, n) = O(\mathrm{polylog}\, n). \end{aligned} \tag{86}$$

Combining (84)–(86), we get

$$\begin{aligned} S_1 &\leq \sigma^{-1}\sum_{\alpha\in\mathcal{A}}\sum_{\beta:\alpha\cap\beta\neq\emptyset}\mu_\alpha \mathrm{E}\left[|Y_\beta V_{\alpha\beta}|\right] \overset{(56)}{\leq} \sigma^{-2}\sum_{\alpha\in\mathcal{A}}\mu_\alpha\left[T_2(\alpha) + \sum_{\beta:\alpha\cap\beta\neq\emptyset} T_1(\alpha,\beta)\right] \\ &\leq O(\sigma^{-2}\cdot \mathrm{polylog}\, n)\left[\mathrm{E}(Y) + k^2\sum_{\beta:\alpha\cap\beta\neq\emptyset}\mu_\beta\right] \overset{\mathbf{Y1}}{\leq} O(\sigma^{-2}\mathrm{E}(Y)\cdot \mathrm{polylog}\, n) \end{aligned}$$

To bound S_2, let $\alpha, \beta \in \mathcal{A}$ be disjoint. As $X_\alpha \le (Y_\alpha + \mu_\alpha)/\sigma$, we obtain

$$\begin{aligned}
\mathrm{E}\left[\left|X_\alpha (Y_\beta - Y_\beta^\alpha) V_{\alpha\beta}\right|\right] &\le \sigma^{-1}\mathrm{E}\left[\left|(Y_\alpha + \mu_\alpha)(Y_\beta - Y_\beta^\alpha) V_{\alpha\beta}\right|\right] \\
&\overset{(56),(60)}{\le} \sigma^{-2}\mathrm{E}\left[\left|(Y_\alpha + \mu_\alpha)(Y_\beta - Y_\beta^\alpha) Y_\beta^\alpha\right|\right] \\
&\qquad + \sigma^{-2} \sum_{\substack{\gamma:\beta\cap\gamma=\emptyset \\ \wedge \alpha\cap\gamma=\emptyset}} \mathrm{E}\left[\left|(Y_\alpha + \mu_\alpha)(Y_\beta - Y_\beta^\alpha)(Y_\gamma^\alpha - Y_\gamma^{\alpha\cup\beta})\right|\right] \\
&\le \sigma^{-2}(T_1 + T_2 + T_3 + T_4),
\end{aligned}$$

where

$$T_1 = \mathrm{E}\left[\left|Y_\alpha (Y_\beta - Y_\beta^\alpha) Y_\beta^\alpha\right|\right],\ T_2 = \mu_\alpha \mathrm{E}\left[\left|(Y_\beta - Y_\beta^\alpha) Y_\beta^\alpha\right|\right],$$
$$T_3 = \sum_{\substack{\gamma:\beta\cap\gamma=\emptyset \\ \wedge \alpha\cap\gamma=\emptyset}} \mathrm{E}\left[\left|Y_\alpha (Y_\beta - Y_\beta^\alpha)(Y_\gamma^\alpha - Y_\gamma^{\alpha\cup\beta})\right|\right],\ T_4 = \mu_\alpha \sum_{\substack{\gamma:\beta\cap\gamma=\emptyset \\ \wedge \alpha\cap\gamma=\emptyset}} \mathrm{E}\left[\left|(Y_\beta - Y_\beta^\alpha)(Y_\gamma^\alpha - Y_\gamma^{\alpha\cup\beta})\right|\right].$$

Now, $T_1 = 0$ by (58). Moreover, bounding T_2 by (62), T_3 by (66) and T_4 by (65), we obtain

$$\begin{aligned}
\sigma^2 \mathrm{E}\left[\left|X_\alpha (Y_\beta - Y_\beta^\alpha) V_{\alpha\beta}\right|\right] &\le O(\frac{\mu_\alpha \mu_\beta}{n} \cdot \operatorname{polylog} n) + \sum_{\substack{\gamma:\beta\cap\gamma=\emptyset \\ \wedge \alpha\cap\gamma=\emptyset}} O(\frac{\mu_\alpha \mu_\beta \mu_\gamma}{n^2} \cdot \operatorname{polylog} n) \\
&= O(\frac{\mu_\alpha \mu_\beta}{n} \cdot \operatorname{polylog} n).
\end{aligned}$$

Thus, (87) yields $S_2 \le \sigma^{-2} \sum_{\beta:\alpha\cap\beta=\emptyset} O(\frac{\mu_\alpha\mu_\beta}{n} \cdot \operatorname{polylog} n) = O(n^{-1}\sigma^{-2}\mathrm{E}(Y)^2 \cdot \operatorname{polylog} n) = O(\sigma^{-2}\mathrm{E}(Y) \cdot \operatorname{polylog} n)$, as desired. □

Lemma 32. $\sum_{\alpha\in\mathcal{A}} \sum_{\beta\in\mathcal{A}} \mathrm{E}\left[|X_\alpha Z_{\alpha\beta}|\right] \mathrm{E}\left[|Z_\alpha + V_{\alpha\beta}|\right] = O(\sigma^{-3}\mathrm{E}\left[Y\right] \cdot \operatorname{polylog} n)$

Proof. Since $|\sigma X_\alpha| \le Y_\alpha + \mu_\alpha$,

$$\begin{aligned}
\sum_{\alpha\in\mathcal{A}} \sum_{\beta\in\mathcal{A}} \mathrm{E}\left[|X_\alpha Z_{\alpha\beta}|\right] \mathrm{E}\left[|Z_\alpha + V_{\alpha\beta}|\right] \le \sigma^{-1}\Big(&\sum_{\alpha\in\mathcal{A}} \sum_{\beta\in\mathcal{A}} \mu_\alpha \mathrm{E}\left[|Z_{\alpha\beta}|\right](\mathrm{E}\left[|Z_\alpha|\right] + \mathrm{E}\left[|V_{\alpha\beta}|\right]) + \\
&\mathrm{E}\left[Y_\alpha \left|Z_{\alpha\beta}\right|\right](\mathrm{E}\left[|Z_\alpha|\right] + \mathrm{E}\left[|V_{\alpha\beta}|\right])\Big). \quad (87)
\end{aligned}$$

Furthermore, we have the three estimates

$$\begin{aligned}
\sigma \mathrm{E}\left[|Z_\alpha|\right] &\le \sigma \sum_{\beta\in\mathcal{A}} \mathrm{E}\left[|Z_{\alpha\beta}|\right] \overset{(55)}{=} \sum_{\beta:\alpha\cap\beta\ne\emptyset} \mu_\beta + \sum_{\beta:\alpha\cap\beta=\emptyset} \mathrm{E}\left[\left|Y_\beta - Y_\beta^\alpha\right|\right] \\
&\overset{(62),\,\mathbf{Y1}}{\le} \sum_{\beta\in\mathcal{A}} O(n^{-1}\mu_\beta \cdot \operatorname{polylog} n) = O(\operatorname{polylog} n), && (88) \\
\sigma \mathrm{E}\left[|V_{\alpha\beta}|\right] &\overset{(56)}{\le} \sum_{\substack{\gamma:\beta\cap\gamma\ne\emptyset \\ \wedge \alpha\cap\gamma=\emptyset}} \mathrm{E}\left[\left|Y_\gamma^\alpha\right|\right] + \sum_{\substack{\gamma:\beta\cap\gamma=\emptyset \\ \wedge \alpha\cap\gamma=\emptyset}} \mathrm{E}\left[\left|Y_\gamma^\alpha - Y_\gamma^{\alpha\cup\beta}\right|\right] \\
&\overset{(69),\,\mathbf{Y1}}{=} \sum_{\gamma\in\mathcal{A}} O(n^{-1}\mu_\gamma \cdot \operatorname{polylog} n) \le O(\operatorname{polylog} n), && (89) \\
\sum_{\beta\in\mathcal{A}} \sigma \mathrm{E}\left[Y_\alpha \left|Z_{\alpha\beta}\right|\right] &\overset{(55)}{=} \sum_{\beta:\alpha\cap\beta\ne\emptyset} \mathrm{E}\left[Y_\alpha Y_\beta\right] + \sum_{\beta:\alpha\cap\beta=\emptyset} \mathrm{E}\left[Y_\alpha \left|Y_\beta - Y_\beta^\alpha\right|\right] \\
&\overset{(59),(63)}{=} k\mu_\alpha + \sum_{\beta:\alpha\cap\beta=\emptyset} \frac{\mu_\alpha \mu_\beta}{n} = O(\mu_\alpha \cdot \operatorname{polylog} n). && (90)
\end{aligned}$$

Now, (88)–(90) yield

$$\sum_{\alpha\in\mathcal{A}}\sum_{\beta\in\mathcal{A}}\mu_\alpha \mathrm{E}\left[|Z_{\alpha\beta}|\right](\mathrm{E}\left[|Z_\alpha|\right]+\mathrm{E}\left[|V_{\alpha\beta}|\right]) = O(\sigma^{-2}\cdot\mathrm{polylog}\,n)\sum_{\alpha\in\mathcal{A}}\mu_\alpha$$
$$= O(\sigma^{-2}\mathrm{E}\left[Y\right]\cdot\mathrm{polylog}\,n), \tag{91}$$
$$\sum_{\alpha\in\mathcal{A}}\sum_{\beta\in\mathcal{A}}\mathrm{E}\left[Y_\alpha\,|Z_{\alpha\beta}|\right](\mathrm{E}\left[|Z_\alpha|\right]+\mathrm{E}\left[|V_{\alpha\beta}|\right]) = O(\sigma^{-2}\cdot\mathrm{polylog}\,n)\sum_{\alpha\in\mathcal{A}}\mu_\alpha$$
$$= O(\sigma^{-2}\mathrm{E}\left[Y\right]\cdot\mathrm{polylog}\,n). \tag{92}$$

Combining (87), (91), and (92), we obtain the assertion. □

Finally, Lemma 27 is an immediate consequence of Lemmas 30–32.

6 Conclusion

Using a purely probabilistic approach, we have established a local limit theorem for $\mathcal{N}(H_d(n,p))$. This result has a number of interesting consequences, which we derive in a follow-up paper [4]. Namely, via Fourier analysis the *univariate* local limit theorem (Theorem 2) can be transformed into a *bivariate* one that describes the joint distribution of the order and the number of edges of the largest component. Furthermore, since given its number of vertices and edges the largest component is a uniformly distributed connected graph, this bivariate limit theorem yields an asymptotic formula for the number of connected hypergraphs with a given number of vertices and edges. Thus, we can solved an involved enumerative problem ("how many connected hypergraphs with ν vertices and μ edges exist?") via a purely probabilistic approach.

The techniques that we have presented in the present paper appear rather generic and may apply to further related problems. For instance, it seems possible to extend our proof of Theorem 2 to the regime $c = \binom{n-1}{d-1}p = (d-1)^{-1} + o(1)$. In addition, it would be interesting to see whether our techniques can be used to obtain limit theorems for the k-core of a random graph, or for the largest component of a random digraph.

References

1. Andriamampianina, T., Ravelomanana, V.: Enumeration of connected uniform hypergraphs. Proceedings of FP-SAC 2005
2. Barraez, D., Boucheron, S., Fernandez de la Vega, W.: On the fluctuations of the giant component. Combinatorics, Probability and Computing **9** (2000) 287–304
3. Barbour, A.D., Karoński, M., Ruciński, A.: A central limit theorem for decomposable random variables with applications to random graphs. J. Combin. Theory Ser. B **47** (1989) 125–145
4. Behrisch, M., Coja-Oghlan, A., Kang, M.: Local limit theorems and the number of connected hypergraphs. Preprint (2007).
5. Bender, E.A., Canfield, E.R., McKay, B.D.: The asymptotic number of labeled connected graphs with a given number of vertices and edges. Random Structures and Algorithms **1** (1990) 127–169
6. Bollobás, B.: Random graphs. 2nd edition. Cambridge University Press (2001)
7. Coja-Oghlan, A., Moore, C., Sanwalani, V.: Counting connected graphs and hypergraphs via the probabilistic method. To appear in Random Structures and Algorithms.
8. Erdős, P., Rényi, A.: On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci. **5** (1960) 17–61.
9. van der Hofstad, R., Spencer, J.: Counting connected graphs asymptotically. To appear in the European Journal on Combinatorics.
10. Janson, S.: The minimal spanning tree in a complete graph and a functional limit theorem for trees in a random graph. Random Structures and Algorithms **7** (1995) 337–355
11. Janson, S., Łuczak, T, Ruciński, A.: Random Graphs, Wiley 2000
12. Karoński, M., Łuczak, T.: The number of connected sparsely edged uniform hypergraphs. Discrete Math. **171** (1997) 153–168
13. Karoński, M., Łuczak, T.: The phase transition in a random hypergraph. J. Comput. Appl. Math. **142** (2002) 125–135

14. Pittel, B.: On tree census and the giant component in sparse random graphs. Random Structures and Algorithms **1** (1990) 311–342
15. Pittel, B., Wormald, N.C.: Counting connected graphs inside out. J. Combin. Theory, Series B **93** (2005) 127–172
16. Ravelomanana, V., Rijamamy, A.L.: Creation and growth of components in a random hypergraph process. Preprint (2005).
17. Schmidt-Pruzan, J., Shamir, E.: Component structure in the evolution of random hypergraphs. Combinatorica **5** (1985) 81–94
18. Stein, C.: A bound for the error in the normal approximation to the distribution of a sum of dependent variables. Proc. 6th Berkeley Symposium on Mathematical Statistics and Probability (1970) 583–602
19. Stepanov, V. E.: On the probability of connectedness of a random graph $\mathcal{G}_m(t)$. Theory Prob. Appl. **15** (1970) 55–67.